

Package ‘reutils’

July 2, 2014

Type Package

Title Talk to the NCBI EUtils

Version 0.1.2

Date 2014-02-21

Author 'Gerhard Schöfl' [aut, cre]

Maintainer 'Gerhard Schöfl' <gschofl@yahoo.de>

Description This is an R package for interfacing with NCBI databases such as PubMed, Genbank, or GEO via the Entrez Programming Utilities (EUtils). The nine EUtils provide a programmatical interface to the NCBI Entrez query and database system for searching and retrieving biological data.

Depends R (>= 3.0.0), methods

Imports assertthat, RCurl, XML

Suggests testthat (>= 0.8)

License MIT + file LICENSE

Encoding UTF-8

URL <https://github.com/gschofl/reutils>

BugReports <https://github.com/gschofl/reutils/issues>

Collate 'utils.R' 'util-error.R' 'util.R' 'ecitmatch.R'
'parse-params.R' 'efetch.R' 'egquery.R' 'einfo.R' 'elink.R'
'epost.R' 'esearch.R' 'espell.R' 'esummary.R' 'parse-docsum.R' 'reutils-package.R' 'zzz.R'

NeedsCompilation no

Repository CRAN

Date/Publication 2014-02-26 17:16:58

R topics documented:

content,ecitmatch-method	2
database,entrez_linkset-method	3
ecitmatch	4
efetch	5
egquery	7
einfo	8
elink	9
epost	11
esearch	12
espell	13
esummary	14
getError	15
getUrl	16
linkset	17
querykey,epost-method	18
retmode	19
rettype	19
reutils	20
uid,entrez_linkset-method	23
webenv,epost-method	24
[,efetch,character-method	25
[,einfo,ANY-method	25
[,elink,numeric-method	26
[,esearch,numeric-method	27
[,esummary,character-method	28
Index	29

content,ecitmatch-method

Extract the data content from an Entrez request

Description

There are four ways to access data returned by an Entrez request: as a character string (`as = "text"`), as a `textConnection` (`as = "textConnection"`), as a parsed XML tree (`as = "xml"`), or, if supported, parsed into a native R object, e.g. a list or a `data.frame` (`as = "parsed"`).

Usage

```
## S4 method for signature 'ecitmatch'
content(x, as = "text", ...)

## S4 method for signature 'efetch'
content(x, as = NULL, ...)
```

```
content(x, as = "xml", ...)  
  
## S4 method for signature 'eutil'  
content(x, as = "xml", ...)
```

Arguments

x	An eutil object.
as	Type of output: "xml", "text", "textConnection", or "parsed".
...	Further arguments passed on to methods.

See Also

[einfo](#), [esearch](#), [esummary](#), [efetch](#), [elink](#), [epost](#), [egquery](#), [espell](#), [ecitmatch](#).

Examples

```
## Not run:  
e <- einfo()  
  
## return XML as an 'XMLInternalDocument'.  
content(e, "xml")  
  
## return XML as character string.  
cat(content(e, "text"))  
  
## return DbNames parsed into a character vector.  
content(e, "parsed")  
  
## return a textConnection to allow linewise read of the data.  
x <- efetch("CP000828", "nucore", rettype = "gbwithparts", retmode = "text")  
con <- content(x, "textConnection")  
readLines(con, 2)  
close(con)  
  
## End(Not run)
```

database,entrez_linkset-method
database

Description

Retrieve the target database name from an [eutil](#) object.

Usage

```
## S4 method for signature 'entrez_linkset'  
database(x, ...)  
  
## S4 method for signature 'entrez_uid'  
database(x, ...)  
  
database(x, ...)  
  
## S4 method for signature 'eutil'  
database(x, ...)
```

Arguments

x An *eutil* object.
... Further arguments passed on to methods.

Value

A character string.

See Also

[einfo](#), [esearch](#), [esummary](#), [efetch](#), [elink](#), [epost](#), [egquery](#), [espell](#), [ecitmatch](#).

Examples

```
## Not run:  
e <- esearch("Mus musculus", "taxonomy")  
database(e)  
  
## End(Not run)
```

ecitmatch

ecitmatch - searching PubMed with citation data

Description

ecitmatch serves as an API to the PubMed [batch citation matcher](#). It retrieves PubMed IDs (PMIDs) that correspond to a set of input citation strings.

Usage

```
ecitmatch(bdata, db = "pubmed", retmode = "xml")
```

Arguments

bdata	Citation strings. Each input citation must be represented by a citation string in the following format: <i>journal_title</i> <i>year</i> <i>volume</i> <i>first_page</i> <i>author_name</i> <i>your_key</i>
db	Database to search. The only supported value is 'pubmed'.
retmode	Retrieval mode The only supported value is 'xml'.

Value

An `ecitmatch` object.

Examples

```
citstrings <- c("proc natl acad sci u s a|1991|88|3248|mann bj|Art1|",
               "science|1987|235|182|palmenber ac|Art2|")
x <- ecitmatch(citstrings)
x
content(x)
```

efetch *efetch - downloading full records*

Description

efetch performs calls to the NCBI EFetch utility to retrieve data records in the requested format for an NCBI Accession Number, one or more primary UIDs, or for a set of UIDs stored in the user's web environment.

Usage

```
efetch(uid, db = NULL, rettype = NULL, retmode = NULL, retstart = NULL,
       retmax = NULL, querykey = NULL, webenv = NULL, strand = NULL,
       seqstart = NULL, seqstop = NULL, complexity = NULL)
```

Arguments

uid	(Required) A list of UIDs provided either as a character vector, as an <code>esearch</code> object, or by reference to a web environment and a query key obtained directly from previous calls to <code>esearch</code> (if <code>usehistory=TRUE</code>), <code>epost</code> or <code>elink</code> . If UIDs are provided as a plain character vector, <code>db</code> must be specified explicitly, and all of the UIDs must be from the database specified by <code>db</code> .
db	(Required if <code>uid</code> is a character vector of UIDs) Database from which to retrieve records. See here for the supported databases.
rettype	A character string specifying the retrieval type, such as 'abstract' or 'medline' for PubMed, 'gp' or 'fasta' for Protein, or 'gb', or 'fasta' for Nuccore. See here for the available values for each database.

retmode	A character string specifying the data mode of the records returned, such as 'text' or 'xml'. See here for the available values for each database.
retstart	Numeric index of the first record to be retrieved.
retmax	Total number of records from the input set to be retrieved.
querykey	An integer specifying which of the UID lists attached to a user's Web Environment will be used as input to efetch. (Usually obtained directly from objects returned by a previous call to esearch , epost or elink .)
webenv	A character string specifying the Web Environment that contains the UID list. (Usually obtained directly from objects returned by a previous call to esearch , epost or elink .)
strand	Strand of DNA to retrieve. (1: plus strand, 2: minus strand)
seqstart	First sequence base to retrieve.
seqstop	Last sequence base to retrieve.
complexity	Data content to return. (0: entire data structure, 1: bioseq, 2: minimal bioseq-set, 3: minimal nuc-prot, 4: minimal pub-set)

Details

See the official online documentation for NCBI's [EUtilities](#) for additional information.

See [here](#) for the default values for rettype and rretmode, as well as a list of the available databases for the EFetch utility.

Value

An [efetch](#) object.

See Also

[content](#), [getUrl](#), [getError](#), [database](#), [retmode](#), [rettype](#).

Examples

```
## Not run:
## From Protein, retrieve a raw GenPept record and write it to a file.
p <- efetch("195055", "protein", "gp")
p

write(content(p, "text"), file = "~/AAD15290.gp")

## Get accessions for a list of GenBank IDs (GIs)
acc <- efetch(c("1621261", "89318838", "68536103", "20807972", "730439"),
             "protein", rettype = "acc")
acc
acc <- strsplit(content(acc), "\n")[[1]]
acc

## Get GIs from a list of accession numbers
gi <- efetch(c("CAB02640.1", "EAS10332.1", "YP_250808.1", "NP_623143.1", "P41007.1"),
```

```
      "protein", "uilst")
gi

## we can conveniently extract the UIDs using the eutil method #xmlValue(xpath)
gi$xmlValue("/IdList/Id")

## or we can extract the contents of the efetch query using the fuction content()
## and use the XML package to retrieve the UIDs
doc <- content(gi)
XML::xpathSApply(doc, "/IdList/Id", XML::xmlValue)

## Get the scientific name for an organism starting with the NCBI taxon id.
tx <- efetch("527031", "taxonomy")
tx

## Convenience accessor for XML nodes of interest using XPath
## Extract the TaxIds of the Lineage
tx["//LineageEx/Taxon/TaxId"]

## Use an XPath expression to extract the scientific name.
tx$xmlValue("/TaxaSet/Taxon/ScientificName")

## End(Not run)
```

egquery

egquery - performing a global Entrez search

Description

egquery retrieves the number of records in all Entrez databases for a single text query.

Usage

```
egquery(term)
```

Arguments

term A valid Entrez text query.

Details

See the official online documentation for NCBI's [EUtilities](#) for additional information.

Value

An [egquery](#) object.

Examples

```
## Determine the number of records for mouse in Entrez.  
e <- egquery("mouse[orgn]")  
e
```

einfo

einfo - getting database statistics and search fields

Description

einfo performs calls to the NCBI EInfo utility to retrieve the names of all valid Entrez databases, or, if db is provided, to retrieve statistics for a single database, including lists of indexing fields and available link names.

Usage

```
einfo(db = NULL)
```

Arguments

db A valid NCBI database name. If NULL, a list of all current NCBI databases is returned.

Details

See the official online documentation for NCBI's [EUtilities](#) for additional information.

Value

A [einfo](#) object.

See Also

[content](#), [getUrl](#), [getError](#).

Examples

```
## Return a list of all current Entrez database names  
einfo()  
  
## Not run:  
## Return statistics for the Entrez Gene database and parse  
## the returned data into a data.frame  
x <- einfo("gene")  
content(x, "parsed")  
  
## End(Not run)
```

 elink

elink - finding related data through Entrez links

Description

elink generates a list of UIDs in a specified Entrez database that are linked to a set of input UIDs in either the same or another database. For instance, the ELink utility can find Entrez gene records linked to records in Entrez Protein.

Usage

```
elink(uid, dbFrom = NULL, dbTo = NULL, linkname = NULL,
      usehistory = FALSE, cmd = "neighbor", correspondence = FALSE,
      querykey = NULL, webenv = NULL, term = NULL, holding = NULL,
      datatype = NULL, reldate = NULL, mindate = NULL, maxdate = NULL)
```

Arguments

uid	(Required) A character vector of UIDs.
dbFrom	Initial database containing the UIDs in the input list.
dbTo	Destination database from which to retrieve linked UIDs. If not provided links will be sought in the database containing the input UIDs.
linkname	Name of the Entrez link to retrieve. Every link in Entrez is given a name of the form <i>dbFrom_dbTo_subset</i> .
usehistory	If TRUE search results are stored directly in the user's Web environment so that they can be used in subsequent calls to esummary or efetch .
cmd	ELink command mode (default: 'neighbor'). See Details.
correspondence	if TRUE correspondence between query UIDs and destination UIDs is preserved.
querykey	Query key.
webenv	Web Environment.
term	Search query to limit the output set of linked UIDs.
holding	Name of LinkOut provider.
datatype	Type of date to limit the search. One of 'mdat' (modification date), 'pdat' (publication date) or 'edat' (Entrez date).
reldate	umber of days back for which search items are returned.
mindate	Minimum date of search range. Format YYYY/MM/DD.
maxdate	Maximum date of search range. Format YYYY/MM/DD.

Details

See the official online documentation for NCBI's [EUtilities](#) for additional information.

If dbTo and dbFrom are set to the same database, ELink will return neighbors within that database.

Elink commands (cmd) specify the function that elink will perform. Available commands are:

- **"neighbor"** (Default) ELink returns a set of UIDs in dbTo linked to the input UIDs in dbFrom.
- **"neighbor_score"** ELink returns a set of UIDs within the same database as the input UIDs along with similarity scores.
- **"neighbor_history"** ELink posts the output UIDs to the Entrez History server and returns a query_key and WebEnv parameter. Alternatively this is achieved by setting usehistory=TRUE

Value

An [elink](#) object.

See Also

Combine calls to ELink with other EUtils: [esummary](#), [efetch](#).

Accessor methods: [content](#), [getUrl](#), [getError](#), [database](#), [uid](#), [linkset](#),

Examples

```
## Find one set of Gene IDs linked to nuccore GIs 34577062 and 24475906
e <- elink(c("34577062", "24475906"), dbFrom = "nuccore", dbTo = "gene")
e

## Not run:
## Find related articles to PMID 20210808
p <- elink("20210808", dbFrom = "pubmed", dbTo = "pubmed")
p

## Extract linked UIDs from the "pubmed" to "pubmed_reviews" link
linkset(p, "pubmed_pubmed_reviews")

## or
p["pubmed_pubmed_reviews"]

## retrieve the abstracts for the first five linked reviews
abstracts <- efetch(p["pubmed_pubmed_reviews"][1:5], rettype = "abstract")

## End(Not run)
```

epost

epost - uploading UIDs to Entrez

Description

epost uses the Entrez EPost utility to upload primary UIDs to the Entrez History server or append a list of UIDs to an existing set of UIDs attached to a Web Environment.

Usage

```
epost(uid, db = NULL, webenv = NULL)
```

Arguments

uid	(Required) List of UIDs provided as a character or as an esearch object.
db	(Required if uid is a character vector) Database containing the UIDs in the input list.
webenv	(Optional) Web Environment. If provided, this parameter specifies the Web Environment that will receive the UIDs sent by epost. epost will create a new query key associated with that Web Environment. The webenv value is usually returned by a previous call to esearch , epost or elink . If no webenv parameter is provided, the EPost utility will create a new Web Environment and post the UIDs to query key 1.

Details

epost returns an integer label called a query key and an encoded cookie string called a Web environment. epost objects can then be used instead of a UID list in subsequent calls to [esummary](#), [efetch](#), or [elink](#).

See the official online documentation for NCBI's [EUtilities](#) for additional information.

Value

An [epost](#) object.

Examples

```
## post a list of protein GIs to the Entrez History server
gi <- c("194680922", "50978626", "28558982", "9507199", "6678417")
p <- epost(gi, "protein")
p
```

 esearch

esearch - searching an Entrez database

Description

esearch performs searches using the the NCBI ESearch utility to retrieve primary UIDs matching a text query. These UIDs can be used in subsequent calls to [esummary](#), [efetch](#), or [elink](#).

Usage

```
esearch(term, db = "nucore", rettype = "uilst", retstart = 0,
        retmax = 100, usehistory = FALSE, webenv = NULL, querykey = NULL,
        field = NULL, datetype = NULL, reldate = NULL, mindate = NULL,
        maxdate = NULL)
```

Arguments

term	A valid Entrez text query.
db	Database to search (default: nucore).
rettype	Retrieval type. (default: 'uilst', alternative: 'count'.)
retstart	Numeric index of the first UID in the retrieved set to be shown in the XML output (default: 0).
retmax	Total number of UIDs to be retrieved (default: 100).
usehistory	If TRUE, search results are posted directly to the Entrez History Server so that they can be used in subsequent calls to esummary , efetch , or elink . Also, usehistory must be set to TRUE for esearch to interpret query key values included in term or to accept a webenv as input.
webenv	Web environment string returned by a previous call to esearch , epost or elink . When provided, esearch will append the results of the search to the pre-existing Web environment. Providing webenv also allows query keys to be used in term so that previous search sets can be combined or limited.
querykey	query key returned by a previous call to esearch , epost or elink . When provided, esearch will find the intersection of the set specified by querykey and the set retrieved by the query in term (i.e. joins the two with AND).
field	Optional. Search field used to limit the entire search term.
datetype	Optional. Type of date to limit the search. One of "mdat" (modification date), "pdat" (publication date) or "edat" (Entrez date)
reldate	Optional. Number of days back for which search items are returned.
mindate	Optional. Minimum date of search range. Format YYYY/MM/DD, YYYY/MM, or YYYY.
maxdate	Optional. Maximum date of search range. Format YYYY/MM/DD, YYYY/MM, or YYYY.

Details

See the official online documentation for NCBI's [EUtilities](#) for additional information on this EUtility.

Value

An [esearch](#) object.

See Also

Combine calls to ESearch with other EUtils: [esummary](#), [efetch](#), [elink](#).

Accessor methods: [content](#), [getUrl](#), [getError](#), [database](#), [uid](#), [webenv](#), [querykey](#).

Examples

```
## Search PubMed for articles with the term "Chlamydia psittaci" in the
## title that were published in 2013.
pmid <- esearch("Chlamydia psittaci[titl] and 2013[pdat]", "pubmed")
pmid

## Not run:
## Extract the query results either as an XML tree or parsed into
## a character vector
xml <- content(pmid, "xml")
uids <- uid(pmid)

## Alternatively post the UIDs to the History Server.
pmid <- esearch("Chlamydia psittaci[titl] and 2013[pdat]", "pubmed",
               usehistory = TRUE)
pmid

## Associate new search results with the existing search results.
pmid2 <- esearch("Chlamydia psittaci[titl] and 2012[pdat]", "pubmed",
                usehistory = TRUE, webenv = webenv(pmid))
pmid2

## End(Not run)
```

espell

espell - retrieving spelling suggestions

Description

For a text query retrieve an XML containing the original query and spelling suggestions.

Usage

```
espell(term, db = "nucore")
```

Arguments

term	An Entrez text query.
db	An Entrez database.

Value

An [espell](#) object.

Examples

```
###
```

esummary	<i>esummary - downloading Document Summaries</i>
----------	--

Description

esummary performs calls to the NCBI ESummary utility to retrieve document summaries (DocSums) for a list of primary UIDs or for a set of UIDs stored in the user's web environment (using the Entrez History server).

Usage

```
esummary(uid, db = NULL, retstart = 1, retmax = 10000, querykey = NULL,
         webenv = NULL, version = "2.0")
```

Arguments

uid	(Required) List of UIDs provided either as a character vector, as an esearch or elink object, or by reference to a Web Environment and a query key obtained directly from objects returned by previous calls to esearch , epost or elink . If UIDs are provided as a plain character vector, db must be specified explicitly, and all of the UIDs must be from the database specified by db.
db	(Required only when id is a character vector of UIDs) Database from which to retrieve DocSums.
retstart	Numeric index of the first DocSum to be retrieved (default: 1).
retmax	Total number of DocSums from the input set to be retrieved (maximum: 10,000).
querykey	An integer specifying which of the UID lists attached to a user's Web Environment will be used as input to efetch . (Usually obtained directly from objects returned by previous esearch , epost or elink calls.)
webenv	A character string specifying the Web Environment that contains the UID list. (Usually obtained directly from objects returned by previous esearch , epost or elink calls.)
version	If "2.0" esummary will retrieve version 2.0 ESummary XML output.

Details

See the official online documentation for NCBI's [EUtilities](#) for additional information.

Value

An [esummary](#) object.

See Also

[content](#), [getUrl](#), [getError](#), [database](#).

Examples

```
## Retrieve the Document Summary information for a set of
## UIDs from the Gene datanase.
ds <- esummary(c("828392", "790", "470338"), "gene")
ds

## Not run:
## parse the XML into a data frame
df <- content(ds, "parsed")
df

## use XPath expressions to extract nodes of interest
ds['//TaxID']

## End(Not run)
```

getError

getError

Description

Retrieve a http or XML parsing error from an [util](#) object.

Usage

```
getError(x, ...)

## S4 method for signature 'util'
getError(x, ...)
```

Arguments

x An [util](#) object.

... Further arguments passed on to methods.

Value

An [eutil_error](#) object.

See Also

[einfo](#), [esearch](#), [esummary](#), [efetch](#), [elink](#), [epost](#), [egquery](#), [espell](#), [ecitmatch](#).

Examples

```
## Not run:
e <- efetch("Nonsensical_accession_nr", "protein", rettype = "fasta")
getError(e)

## End(Not run)
```

getUrl

getUrl

Description

Retrieve the URL used to perform an Entrez E-Utilities query.

Usage

```
getUrl(x, ...)
```

S4 method for signature 'eutil'

```
getUrl(x, ...)
```

Arguments

x An [eutil](#) object.

... Further arguments passed on to methods.

Value

A character string.

See Also

[einfo](#), [esearch](#), [esummary](#), [efetch](#), [elink](#), [epost](#), [egquery](#), [espell](#), [ecitmatch](#).

Examples

```
## Not run:
e <- efetch("AV333213.1", "protein", rettype = "fasta")
getUrl(e)

## End(Not run)
```

linkset	<i>linkset</i>
---------	----------------

Description

Retrieve a linkset from an [elink](#) object.

Usage

```
linkset(x, linkname = NULL, ...)  
  
## S4 method for signature 'entrez_linkset'  
linkset(x, linkname = NULL, ...)  
  
## S4 method for signature 'elink'  
linkset(x, linkname = NULL, ...)
```

Arguments

x	An elink object.
linkname	(optional) Name of the Entrez link to retrieve. Every link in Entrez is given a name of the form <i>dbFrom_dbTo_subset</i> . If NULL, all available links are retrieved from the object.
...	Further arguments passed on to methods.

Value

A list.

Examples

```
## Not run:  
## Find related articles to PMID 20210808 and extract linked UIDs from the  
## "pubmed" to "pubmed_reviews" link  
x <- elink("20210808", dbFrom = "pubmed", dbTo = "pubmed", cmd = "neighbor_score")  
linkset(x, "pubmed_pubmed_reviews")  
  
## End(Not run)
```

querykey, epost-method *querykey*

Description

An integer query key returned by an ESearch, EPost or ELink call if the History server was used. Otherwise NA.

Usage

```
## S4 method for signature 'epost'  
querykey(x, ...)  
  
## S4 method for signature 'entrez_uid'  
querykey(x, ...)  
  
## S4 method for signature 'esearch'  
querykey(x, ...)  
  
querykey(x, ...)
```

Arguments

x An [eutil](#) object.
... Further arguments passed on to methods.

Value

An integer or NA.

See Also

[einfo](#), [esearch](#), [esummary](#), [efetch](#), [elink](#), [epost](#), [egquery](#), [espell](#), [ecitmatch](#).

Examples

```
## Not run:  
e <- esearch("Mus musculus", "taxonomy", usehistory = TRUE)  
querykey(e)  
  
## End(Not run)
```

retmode	<i>retmode</i>
---------	----------------

Description

Get the “retrieval mode” of an [eutil](#) object. It is usually one of `xml`, `text`, or `asn.1`. It is set to `NULL` if “retrieval mode” is not supported by an E-Utility.

Usage

```
retmode(x, ...)  
  
## S4 method for signature 'eutil'  
retmode(x, ...)
```

Arguments

<code>x</code>	An eutil object.
<code>...</code>	Further arguments passed on to methods.

Value

A character string or `NULL`.

See Also

[einfo](#), [esearch](#), [esummary](#), [efetch](#), [elink](#), [epost](#), [egquery](#), [espell](#), [ecitmatch](#).

Examples

```
## Not run:  
e <- efetch("10090", "taxonomy")  
retmode(e)  
  
## End(Not run)
```

rettype	<i>rettype</i>
---------	----------------

Description

Get the “retrieval type” of an [eutil](#) object. See [here](#) for the available retrieval types for different NCBI databases.

Usage

```
rettype(x, ...)  
  
## S4 method for signature 'eutil'  
rettype(x, ...)
```

Arguments

x An [eutil](#) object.
... Further arguments passed on to methods.

Value

A character string.

See Also

[einfo](#), [esearch](#), [esummary](#), [efetch](#), [elink](#), [epost](#), [egquery](#), [espell](#), [ecitmatch](#).

Examples

```
## Not run:  
e <- esearch("Mus musculus", "taxonomy")  
rettype(e)  
  
## End(Not run)
```

reutils

An interface to the NCBI Entrez programming Utilities.

Description

reutils provides support for interacting with NCBI databases such as PubMed, Genbank, or GEO via the Entrez Programming Utilities ([EUtils](#)).

Please check the relevant [usage guidelines](#) when using these services. Note that Entrez server requests are subject to frequency limits.

Details

With nine E-Utilities, NCBI provides a programmatical interface to the Entrez query and database system for searching and retrieving requested data

Each of these tools corresponds to an R function in the reutils package described below.

The output returned by the EUtils is typically in XML format. To gain access to this output you have several options:

1. Use the `content(as = "xml")` method to extract the output as an `XMLInternalDocument` object and process it further using the facilities provided by the XML package.

2. Use the `content(as = "parsed")` method to extract the output into `data.frames`. Note that this is currently only implemented for `docsums` returned by `esummary`, `uilsts` returned by `esearch`, and the output returned by `einfo`.
3. Access specific nodes in the XML tree using XPath expressions with the reference class methods `#xmlValue`, `#xmlAttr`, or `#xmlName` built into `util` objects.

The Entrez Programming Utilities can also generate output in other formats, such as plain-text Fasta or GenBank files for sequence databases, or the MedLine format for the literature database. The type of output is generally controlled by setting the `retmode` and `rettype` arguments when calling a `EUtil`.

Main functions

- `esearch`: Search and retrieve primary UIDs for use with `esummary`, `elink`, or `efetch`. `esearch` additionally returns term translations and optionally stores results for future use in the user's Web Environment.
- `esummary`: Retrieve document summaries from a list of primary UIDs (Provided as a character vector or as an `esearch` object).
- `egquery`: Provides Entrez database counts in XML for a single search term using a Global Query.
- `einfo`: Retrieve field names, term counts, last update, and available updates for each database.
- `efetch`: Retrieve data records in a specified format corresponding to a list of primary UIDs or from the user's Web Environment in the Entrez History server.
- `elink`: Returns a list of UIDs (and relevancy scores) from a target database that are related to a list of UIDs in the same database or in another Entrez database.
- `epost`: Uploads primary UIDs to the user's Web Environment on the Entrez history server for subsequent use with `esummary`, `elink`, or `efetch`.
- `espell`: Provide spelling suggestions.
- `ecitmatch`: Retrieves PubMed IDs (PMIDs) that correspond to a set of input citation strings
- `content`: Extract the content of a request from the `util` object returned by any of the above functions.

Package options

`reutils` uses three `options` to configure behaviour:

- `reutils.email`: NCBI requires that a user of their API provides an email address with a call to Entrez. If you are going to perform a lot of queries consider setting `reutils.email` to your email address in your `.Rprofile` file.
- `reutils.show.headlines`: By default `efetch` objects containing text data show only the first 12 lines. This is quite handy if you have downloaded a fairly large genome in Genbank file format. This can be changed by setting the global option `reutils.show.headlines` to another numeric value or `NULL`.
- `reutils.verbose.queries`: If you perform many queries interactively you might want to get messages announcing the queries you run. You can do so by setting the option `reutils.verbose.queries` to `TRUE`.

- `reutils.test.remote`: Unit tests that require online access to NCBI services are disabled by default, as they cannot be guaranteed to be available/working under all circumstances. Set the option `codereutils.test.remote` to `TRUE` to run the full suite of tests.

Author(s)

Gerhard Schöfl <gschofl@yahoo.de>

Examples

```
#
# combine esearch and efetch
#
# Download PubMed records that are indexed in MeSH for both 'Chlamydia' and
# 'genome' and were published in 2013.
query <- "Chlamydia[mesh] and genome[mesh] and 2013[pdat]"

# Upload the PMIDs for this search to the History server
pmids <- esearch(query, "pubmed", usehistory = TRUE)
pmids

## Not run:
# Fetch the records
articles <- efetch(pmids)

# Use XPath expressions with the #xmlValue() or #xmlAttr() methods to directly
# extract specific data from the XML records stored in the 'efetch' object.
titles <- articles$xmlValue("//ArticleTitle")
abstracts <- articles$xmlValue("//AbstractText")

#
# combine epost with esummary/efetch
#
# Download protein records corresponding to a list of GI numbers.
uid <- c("194680922", "50978626", "28558982", "9507199", "6678417")

# post the GI numbers to the Entrez history server
p <- epost(uid, "protein")

# retrieve docsums with esummary
docsum <- content(esummary(p, version = "1.0"), "parsed")
docsum

# download FASTAs as 'text' with efetch
prot <- efetch(p, retmode = "text", rettype = "fasta")
prot

# retrieve the content from the efetch object
fasta <- content(prot)

## End(Not run)
```

```
uid,entrez_linkset-method
      uid
```

Description

Retrieve the list of UIDs returned by a call to ESearch or ELink.

Usage

```
## S4 method for signature 'entrez_linkset'
uid(x, ...)

## S4 method for signature 'elink'
uid(x, ...)

## S4 method for signature 'entrez_uid'
uid(x, ...)

## S4 method for signature 'esearch'
uid(x, ...)

uid(x, ...)
```

Arguments

x	An eutil object.
...	Further arguments passed on to methods.

Value

A character vector.

See Also

[esearch](#), [elink](#).

Examples

```
## Not run:
e <- esearch("Mus musculus", "taxonomy")
uid(e)

## End(Not run)
```

webenv, epost-method *webenv*

Description

Retrieve the Web environment string returned from an ESearch, EPost or ELink call. NA if the History server was not used.

Usage

```
## S4 method for signature 'epost'  
webenv(x, ...)  
  
## S4 method for signature 'entrez_uid'  
webenv(x, ...)  
  
## S4 method for signature 'esearch'  
webenv(x, ...)  
  
webenv(x, ...)
```

Arguments

x An [eutil](#) object.
... Further arguments passed on to methods.

Value

A character string or NA.

See Also

[einfo](#), [esearch](#), [esummary](#), [efetch](#), [elink](#), [epost](#), [egquery](#), [espell](#), [ecitmatch](#).

Examples

```
## Not run:  
e <- esearch("Mus musculus", "taxonomy", usehistory = TRUE)  
webenv(e)  
  
## End(Not run)
```

[,efetch,character-method
EFetch accessors

Description

Extract XML nodes from an [efetch](#) object.

Usage

```
## S4 method for signature 'efetch,character'  
x[i]  
  
## S4 method for signature 'efetch,character'  
x[[i]]
```

Arguments

x An [efetch](#) object containing XML data.
i An XPath expression specifying the XML nodes to extract.

Value

An XML node set.

Examples

```
## Not run:  
p <- efetch("195055", "protein", "gp", "xml")  
p['//GBFeature[GBFeature_key="mat_peptide"]//GBQualifier_value']  
  
## End(Not run)
```

[,einfo,ANY-method *EInfo accessors*

Description

Extract parts of a parsed [einfo](#) object.

Usage

```
## S4 method for signature 'einfo,ANY'  
x[i]  
  
## S4 method for signature 'einfo,ANY'  
x[[i]]
```

Arguments

- x An `einfo` object.
- i Numeric or character indices specifying the elements to extract.

Value

A list.

See Also

[Extract](#)

Examples

```
## Not run:
e <- einfo("pubmed")
e[1:5]
e["Description"]
e[["Links"]]

## End(Not run)
```

[,elink,numeric-method

ELink Accessors

Description

Extract UIDs from an `elink` object.

Usage

```
## S4 method for signature 'elink,numeric'
x[i]

## S4 method for signature 'elink,character'
x[i]
```

Arguments

- x An `elink` object.
- i Numeric or character indices.

Value

A `entrez_linkset` object.

Examples

```
## Not run:
e <- elink(c("34577062", "24475906"), dbFrom = "nuccore")
e[1]

## End(Not run)
```

[,esearch,numeric-method
ESearch Accessors

Description

Extract UIDs from an [esearch](#) object.

Usage

```
## S4 method for signature 'esearch,numeric'
x[i]
```

Arguments

x	An esearch object.
i	Numeric indices.

Value

A [entrez_uid](#) object.

Examples

```
## Not run:
e <- esearch("Mus musculus", "protein", retmax = 20)
e[1:5]
## pass the subset directly on to esummary or efetch
content(esummary(e[1:5]), "parsed")

## End(Not run)
```

[,esummary,character-method
ESummary accessors

Description

Extract XML nodes from an [esummary](#) object.

Usage

```
## S4 method for signature 'esummary,character'  
x[i]
```

```
## S4 method for signature 'esummary,character'  
x[[i]]
```

Arguments

x	An esummary object.
i	An XPath expression.

Value

An XML node set.

Examples

```
## Not run:  
ds <- esummary("470338", "protein")  
ds["//Slen/node()"]  
  
require("XML")  
as.numeric(xmlValue(ds["//Slen"]))  
  
## End(Not run)
```

Index

*Topic **package**

- reutils, [20](#)
- [,efetch,character-method, [25](#)
- [,einfo,ANY-method, [25](#)
- [,elink,character-method
 - ([,elink,numeric-method), [26](#)
- [,elink,numeric-method, [26](#)
- [,esearch,numeric-method, [27](#)
- [,esummary,character-method, [28](#)
- [[,efetch,character-method
 - ([,efetch,character-method), [25](#)
- [[,einfo,ANY-method
 - ([,einfo,ANY-method), [25](#)
- [[,esummary,character-method
 - ([,esummary,character-method), [28](#)

- content, [6, 8, 10, 13, 15, 21](#)
- content (content,ecitmatch-method), [2](#)
- content,ecitmatch-method, [2](#)
- content,efetch-method
 - (content,ecitmatch-method), [2](#)
- content,util-method
 - (content,ecitmatch-method), [2](#)

- database, [6, 10, 13, 15](#)
- database
 - (database,entrez_linkset-method), [3](#)
- database,entrez_linkset-method, [3](#)
- database,entrez_uid-method
 - (database,entrez_linkset-method), [3](#)
- database,util-method
 - (database,entrez_linkset-method), [3](#)

- ecitmatch, [3, 4, 4, 5, 16, 18–21, 24](#)
- efetch, [3, 4, 5, 6, 9–13, 16, 18–21, 24, 25](#)
- egquery, [3, 4, 7, 7, 16, 18–21, 24](#)

- einfo, [3, 4, 8, 8, 16, 18–21, 24–26](#)
- elink, [3–6, 9, 10–14, 16–21, 23, 24, 26](#)
- entrez_linkset, [26](#)
- entrez_uid, [27](#)
- epost, [3–6, 11, 11, 12, 14, 16, 18–21, 24](#)
- esearch, [3–6, 11, 12, 12, 13, 14, 16, 18–21, 23, 24, 27](#)
- espell, [3, 4, 13, 14, 16, 18–21, 24](#)
- esummary, [3, 4, 9–13, 14, 15, 16, 18–21, 24, 28](#)
- util, [3, 4, 15, 16, 18–21, 23, 24](#)
- util_error, [16](#)
- Extract, [26](#)

- getError, [6, 8, 10, 13, 15, 15](#)
- getError,util-method (getError), [15](#)
- getUrl, [6, 8, 10, 13, 15, 16](#)
- getUrl,util-method (getUrl), [16](#)

- linkset, [10, 17](#)
- linkset,elink-method (linkset), [17](#)
- linkset,entrez_linkset-method (linkset), [17](#)

- options, [21](#)

- querykey, [13](#)
- querykey (querykey,epost-method), [18](#)
- querykey,entrez_uid-method (querykey,epost-method), [18](#)
- querykey,epost-method, [18](#)
- querykey,esearch-method (querykey,epost-method), [18](#)

- retmode, [6, 19](#)
- retmode,util-method (retmode), [19](#)
- rettype, [6, 19](#)
- rettype,util-method (rettype), [19](#)
- reutils, [20](#)
- reutils-package (reutils), [20](#)

- textConnection, [2](#)

uid, [10](#), [13](#)
uid (uid,entrez_linkset-method), [23](#)
uid,elink-method
 (uid,entrez_linkset-method), [23](#)
uid,entrez_linkset-method, [23](#)
uid,entrez_uid-method
 (uid,entrez_linkset-method), [23](#)
uid,esearch-method
 (uid,entrez_linkset-method), [23](#)

webenv, [13](#)
webenv (webenv,epost-method), [24](#)
webenv,entrez_uid-method
 (webenv,epost-method), [24](#)
webenv,epost-method, [24](#)
webenv,esearch-method
 (webenv,epost-method), [24](#)