Package 'PubChemR'

January 16, 2024

Type Package

Title Interface to the 'PubChem' Database for Chemical Data Retrieval

Version 1.2

Description Provides an interface to the 'PubChem' database via the PUG REST <https://pubchem.ncbi.nlm.nih.gov/docs/pug-rest> and PUG View <https://pubchem.ncbi.nlm.nih.gov/docs/pug-view> services. This package allows users to automatically access chemical and biological data from 'PubChem', including compounds, substances, assays, and various other data types.
Functions are available to retrieve data in different formats, perform searches, and access detailed annotations.

License GPL (>= 2)

Encoding UTF-8

RoxygenNote 7.2.3

Imports dplyr, httr, RJSONIO, RCurl, tibble, tidyr, utils, magrittr, stringr, magick, png, rsvg, xml2, XML

NeedsCompilation no

Author Selcuk Korkmaz [aut, cre] (<https://orcid.org/0000-0003-4632-6850>), Bilge Eren Yamassan [aut] (<https://orcid.org/0000-0002-6525-2503>), Dincer Goksuluk [aut] (<https://orcid.org/0000-0002-2752-7668>)

Maintainer Selcuk Korkmaz <selcukorkmaz@gmail.com>

Repository CRAN

Date/Publication 2024-01-16 11:40:06 UTC

R topics documented:

`api_base` *PubChem API Base URL*

Description

The base URL for the PubChem API.

Usage

```
api_base
```

Format

An object of class `character` of length 1.

`BondType` *Bond Types*

Description

List of bond types.

Usage

```
BondType
```

Format

An object of class `list` of length 8.

`CompoundIdType` *Compound ID Types*

Description

List of compound ID types.

Usage

```
CompoundIdType
```

Format

An object of class `list` of length 8.

`CoordinateType` *Coordinate Types*

Description

List of coordinate types.

Usage

```
CoordinateType
```

Format

An object of class `list` of length 15.

download *Download Content from PubChem and Save to a File*

Description

This function sends a request to PubChem to retrieve content in the specified format for a given identifier. It then writes the content to a specified file path.

Usage

```
download(
  filename = NULL,
  outformat,
  path,
  identifier,
  namespace = "cid",
  domain = "compound",
  operation = NULL,
  searchtype = NULL,
  overwrite = FALSE,
  options = NULL
)
```

Arguments

filename	a character string specifying the file name to be saved. If not specified, a default file name "file" is used.
outformat	A character string specifying the desired output format (e.g., "sdf", "json").
path	A character string specifying the path where the content should be saved.
identifier	A vector of positive integers (e.g. cid, sid, aid) or identifier strings (source, inchikey, formula). In some cases, only a single identifier string (name, smiles, xref; inchi, sdf by POST only).
namespace	Specifies the namespace for the query. For the 'compound' domain, possible values include 'cid', 'name', 'smiles', 'inchi', 'sdf', 'inchikey', 'formula', 'substructure', 'superstructure', 'similarity', 'identity', 'xref', 'listkey', 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure', and 'fastformula'. For other domains, the possible namespaces are domain-specific.
domain	Specifies the domain of the query. Possible values are 'substance', 'compound', 'assay', 'gene', 'protein', 'pathway', 'taxonomy', 'cell', 'sources', 'sourcetable', 'conformers', 'annotations', 'classification', and 'standardize'.
operation	Specifies the operation to be performed on the input records. For the 'compound' domain, possible operations include 'record', 'property', 'synonyms', 'sids', 'cids', 'aids', 'assaysummary', 'classification', 'xrefs', and 'description'. The available operations are domain-specific.

searchtype	Specifies the type of search to be performed. For structure searches, possible values are combinations of 'substructure', 'superstructure', 'similarity', 'identity' with 'smiles', 'inchi', 'sdf', 'cid'. For fast searches, possible values are combinations of 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure' with 'smiles', 'smarts', 'inchi', 'sdf', 'cid', or 'fastformula'.
overwrite	A logical value indicating whether to overwrite the file if it already exists. Default is FALSE.
options	Additional arguments.

Value

No return value. The function writes the content to the specified file path and prints a message indicating the save location.

Examples

```
# Download JSON file for the compound "aspirin" into "Aspirin.JSON"
# A folder named "Compound" will be created under current directory"
download(
  filename = "Aspirin",
  outformat = "json",
  path = "./Compound",
  identifier = "aspirin",
  namespace = "name",
  domain = "compound",
  operation = NULL,
  searchtype = NULL,
  overwrite = TRUE
)

# Remove downloaded files and folders.
file.remove("./Compound/Aspirin.json")
file.remove("./Compound/")
```

ELEMENTS *Elements*

Description

Vector of chemical elements.

Usage

```
ELEMENTS
```

Format

An object of class `character` of length 118.

get_aids *Retrieve Assay IDs (AIDs) from PubChem*

Description

This function sends a request to PubChem to retrieve Assay IDs (AIDs) for a given identifier. It returns either a tibble (data frame) with the provided identifier and the corresponding AIDs or a list of AIDs, depending on the 'as_data_frame' parameter.

Usage

```
get_aids(
  identifier,
  namespace = "cid",
  domain = "compound",
  searchtype = NULL,
  as_data_frame = TRUE,
  options = NULL
)
```

Arguments

identifier	A vector of positive integers (e.g. cid, sid, aid) or identifier strings (source, inchikey, formula). In some cases, only a single identifier string (name, smiles, xref; inchi, sdf by POST only).
namespace	Specifies the namespace for the query. For the 'compound' domain, possible values include 'cid', 'name', 'smiles', 'inchi', 'sdf', 'inchikey', 'formula', 'substructure', 'superstructure', 'similarity', 'identity', 'xref', 'listkey', 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure', and 'fastformula'. For other domains, the possible namespaces are domain-specific.
domain	Specifies the domain of the query. Possible values are 'substance', 'compound', 'assay', 'gene', 'protein', 'pathway', 'taxonomy', 'cell', 'sources', 'sourcetable', 'conformers', 'annotations', 'classification', and 'standardize'.
searchtype	Specifies the type of search to be performed. For structure searches, possible values are combinations of 'substructure', 'superstructure', 'similarity', 'identity' with 'smiles', 'inchi', 'sdf', 'cid'. For fast searches, possible values are combinations of 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure' with 'smiles', 'smarts', 'inchi', 'sdf', 'cid', or 'fastformula'.
as_data_frame	A logical value indicating whether to return the results as a tibble (data frame). Default is TRUE.
options	Additional arguments passed to get_json

Value

If ‘as_data_frame‘ is TRUE, a tibble (data frame) where each row corresponds to a provided identifier and its AID. The tibble has columns ’CID’ and ’AID’. If ‘as_data_frame‘ is FALSE, a list of AIDs is returned.

Examples

```
get_aids(
  identifier = "aspirin",
  namespace = "name"
)
```

---

get_all_sources *Retrieve All Sources from PubChem*

Description

This function retrieves a list of all current depositors of substances or assays from PubChem.

Usage

```
get_all_sources(domain = "substance")
```

Arguments

domain	A character string specifying the domain for which sources (’substance’, ’assay’) are to be retrieved. Default is ’substance’.

Value

A character vector containing the names of all sources for the specified domain.

Examples

```
get_all_sources(
  domain = 'substance'
)
```

get_assays *Retrieve Assays from PubChem*

Description

This function sends a request to PubChem to retrieve assay data based on the specified parameters. It returns a list of assays corresponding to the provided identifiers.

Usage

```
get_assays(
  identifier,
  namespace = "aid",
  operation = NULL,
  searchtype = NULL,
  options = NULL
)
```

Arguments

identifier	A vector of positive integers (e.g. cid, sid, aid) or identifier strings (source, inchikey, formula). In some cases, only a single identifier string (name, smiles, xref; inchi, sdf by POST only).
namespace	Specifies the namespace for the query. For the 'compound' domain, possible values include 'cid', 'name', 'smiles', 'inchi', 'sdf', 'inchikey', 'formula', 'substructure', 'superstructure', 'similarity', 'identity', 'xref', 'listkey', 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure', and 'fastformula'. For other domains, the possible namespaces are domain-specific.
operation	The operation to be performed (default: NULL).
searchtype	The type of search to be performed (default: NULL).
options	Additional parameters. Currently has no effect on the results.

Value

A named list where each element corresponds to an assay retrieved from PubChem. The names of the list elements are based on the provided identifiers. If no assay is found for a given identifier, the corresponding list element will contain the string "No assay".

Examples

```
get_assays(
  identifier = 1234,
  namespace = "aid"
)
```

get_cids *Retrieve Compound IDs (CIDs) from PubChem*

Description

This function sends a request to PubChem to retrieve Compound IDs (CIDs) for a given identifier. It returns a tibble (data frame) with the provided identifier and the corresponding CIDs.

Usage

```
get_cids(
  identifier,
  namespace = "name",
  domain = "compound",
  searchtype = NULL,
  options = NULL
)
```

Arguments

identifier	A vector of positive integers (e.g. cid, sid, aid) or identifier strings (source, inchikey, formula). In some cases, only a single identifier string (name, smiles, xref; inchi, sdf by POST only).
namespace	Specifies the namespace for the query. For the 'compound' domain, possible values include 'cid', 'name', 'smiles', 'inchi', 'sdf', 'inchikey', 'formula', 'substructure', 'superstructure', 'similarity', 'identity', 'xref', 'listkey', 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure', and 'fastformula'. For other domains, the possible namespaces are domain-specific.
domain	Specifies the domain of the query. Possible values are 'substance', 'compound', 'assay', 'gene', 'protein', 'pathway', 'taxonomy', 'cell', 'sources', 'sourcetable', 'conformers', 'annotations', 'classification', and 'standardize'.
searchtype	Specifies the type of search to be performed. For structure searches, possible values are combinations of 'substructure', 'superstructure', 'similarity', 'identity' with 'smiles', 'inchi', 'sdf', 'cid'. For fast searches, possible values are combinations of 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure' with 'smiles', 'smarts', 'inchi', 'sdf', 'cid', or 'fastformula'.
options	Additional arguments passed to get_json.

Value

A tibble (data frame) where each row corresponds to a provided identifier and its CID. The tibble has columns 'Compound' and 'CID'.

Examples

```
get_cids(
  identifier = "aspirin",
  namespace = "name"
)
```

`get_compounds` *Retrieve Compounds from PubChem*

Description

This function sends a request to PubChem to retrieve compound data based on the specified parameters. It returns a list of compounds corresponding to the provided identifiers.

Usage

```
get_compounds(
  identifier,
  namespace = "cid",
  operation = NULL,
  searchtype = NULL,
  options = NULL
)
```

Arguments

`identifier`	A vector of positive integers (e.g. cid, sid, aid) or identifier strings (source, inchikey, formula). In some cases, only a single identifier string (name, smiles, xref; inchi, sdf by POST only).
`namespace`	Specifies the namespace for the query. For the 'compound' domain, possible values include 'cid', 'name', 'smiles', 'inchi', 'sdf', 'inchikey', 'formula', 'substructure', 'superstructure', 'similarity', 'identity', 'xref', 'listkey', 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure', and 'fastformula'. For other domains, the possible namespaces are domain-specific.
`operation`	The operation to be performed (default: NULL).
`searchtype`	Specifies the type of search to be performed. For structure searches, possible values are combinations of 'substructure', 'superstructure', 'similarity', 'identity' with 'smiles', 'inchi', 'sdf', 'cid'. For fast searches, possible values are combinations of 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure' with 'smiles', 'smarts', 'inchi', 'sdf', 'cid', or 'fastformula'.
`options`	Additional parameters passed to `get_json`.

Value

A named list where each element corresponds to a compound retrieved from PubChem. The names of the list elements are based on the provided identifiers. If no compound is found for a given identifier, the corresponding list element will contain the string "No compound".

Examples

```
get_compounds(
  identifier = "aspirin",
  namespace = "name"
)
```

---

get_json *Retrieve JSON Data from PubChem*

Description

This function sends a request to PubChem to retrieve JSON data based on the specified parameters. It handles errors and warnings gracefully, providing informative messages when they occur.

Usage

```
get_json(
  identifier,
  namespace = "cid",
  domain = "compound",
  operation = NULL,
  searchtype = NULL,
  options = NULL
)
```

Arguments

identifier	A vector of positive integers (e.g. cid, sid, aid) or identifier strings (source, inchikey, formula). In some cases, only a single identifier string (name, smiles, xref; inchi, sdf by POST only).
namespace	Specifies the namespace for the query. For the 'compound' domain, possible values include 'cid', 'name', 'smiles', 'inchi', 'sdf', 'inchikey', 'formula', 'substructure', 'superstructure', 'similarity', 'identity', 'xref', 'listkey', 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure', and 'fastformula'. For other domains, the possible namespaces are domain-specific.
domain	Specifies the domain of the query. Possible values are 'substance', 'compound', 'assay', 'gene', 'protein', 'pathway', 'taxonomy', 'cell', 'sources', 'sourcetable', 'conformers', 'annotations', 'classification', and 'standardize'.

operation	Specifies the operation to be performed on the input records. For the 'compound' domain, possible operations include 'record', 'property', 'synonyms', 'sids', 'cids', 'aids', 'assaysummary', 'classification', 'xrefs', and 'description'. The available operations are domain-specific.
searchtype	Specifies the type of search to be performed. For structure searches, possible values are combinations of 'substructure', 'superstructure', 'similarity', 'identity' with 'smiles', 'inchi', 'sdf', 'cid'. For fast searches, possible values are combinations of 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure' with 'smiles', 'smarts', 'inchi', 'sdf', 'cid', or 'fastformula'.
options	Additional parameters passed to get_json.

Value

A list containing the parsed JSON response from PubChem. Returns NULL if an error or warning occurs.

Examples

```
get_json(
  identifier = "aspirin",
  namespace = "name"
)
```

---

get_properties *Retrieve Compound Properties from PubChem*

Description

This function sends a request to PubChem to retrieve compound properties based on the specified parameters. It returns a list or dataframe of properties corresponding to the provided identifiers.

Usage

```
get_properties(
  properties,
  identifier,
  namespace = "cid",
  searchtype = NULL,
  as_dataframe = FALSE,
  options = NULL
)
```

Arguments

properties	A character vector specifying the properties to be retrieved.
identifier	A vector of positive integers (e.g. cid, sid, aid) or identifier strings (source, inchikey, formula). In some cases, only a single identifier string (name, smiles, xref; inchi, sdf by POST only).
namespace	Specifies the namespace for the query. For the 'compound' domain, possible values include 'cid', 'name', 'smiles', 'inchi', 'sdf', 'inchikey', 'formula', 'substructure', 'superstructure', 'similarity', 'identity', 'xref', 'listkey', 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure', and 'fastformula'. For other domains, the possible namespaces are domain-specific.
searchtype	Specifies the type of search to be performed. For structure searches, possible values are combinations of 'substructure', 'superstructure', 'similarity', 'identity' with 'smiles', 'inchi', 'sdf', 'cid'. For fast searches, possible values are combinations of 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure' with 'smiles', 'smarts', 'inchi', 'sdf', 'cid', or 'fastformula'.
as_dataframe	If TRUE it return a dataframe.
options	Additional arguments passed to get_json.

Value

If 'as_dataframe' is FALSE, a named list where each element corresponds to the properties retrieved from PubChem. If 'as_dataframe' is TRUE, a dataframe where each row corresponds to the properties retrieved from PubChem. The names of the list elements or row names of the dataframe are based on the provided identifiers.

Examples

```
get_properties(
  properties = "IsomericSMILES",
  identifier = "aspirin",
  namespace = "name"
)
```

---

get_pubchem *Get Data from PubChem API*

Description

This function constructs a URL to query the PubChem API based on the provided parameters and returns the response content.

Usage

```
get_pubchem(
  identifier,
  namespace = "cid",
  domain = "compound",
  operation = NULL,
  output = "JSON",
  searchtype = NULL,
  options = NULL
)
```

Arguments

identifier	A vector of positive integers (e.g. cid, sid, aid) or identifier strings (source, inchikey, formula). In some cases, only a single identifier string (name, smiles, xref; inchi, sdf by POST only).
namespace	Specifies the namespace for the query. For the 'compound' domain, possible values include 'cid', 'name', 'smiles', 'inchi', 'sdf', 'inchikey', 'formula', 'substructure', 'superstructure', 'similarity', 'identity', 'xref', 'listkey', 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure', and 'fastformula'. For other domains, the possible namespaces are domain-specific.
domain	Specifies the domain of the query. Possible values are 'substance', 'compound', 'assay', 'gene', 'protein', 'pathway', 'taxonomy', 'cell', 'sources', 'sourcetable', 'conformers', 'annotations', 'classification', and 'standardize'.
operation	Specifies the operation to be performed on the input records. For the 'compound' domain, possible operations include 'record', 'property', 'synonyms', 'sids', 'cids', 'aids', 'assaysummary', 'classification', 'xrefs', and 'description'. The available operations are domain-specific.
output	Specifies the desired output format. Possible values are 'XML', 'ASNT', 'ASNB', 'JSON', 'JSONP', 'SDF', 'CSV', 'PNG', and 'TXT'.
searchtype	Specifies the type of search to be performed. For structure searches, possible values are combinations of 'substructure', 'superstructure', 'similarity', 'identity' with 'smiles', 'inchi', 'sdf', 'cid'. For fast searches, possible values are combinations of 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure' with 'smiles', 'smarts', 'inchi', 'sdf', 'cid', or 'fastformula'.
options	Additional arguments passed to request.

Value

Returns the response content from the PubChem API based on the constructed URL.

Examples

```
get_pubchem(
  identifier = "aspirin",
```

```
  namespace = "name"
)
```

`get_pug_rest` *Retrieve Data from PubChem PUG REST API*

Description

This function sends a request to the PubChem PUG REST API to retrieve various types of data for a given identifier. It supports fetching data in different formats and allows saving the output.

Usage

```
get_pug_rest(
  identifier = NULL,
  namespace = "cid",
  domain = "compound",
  operation = NULL,
  output = "JSON",
  searchtype = NULL,
  property = NULL,
  options = NULL,
  saveFile = FALSE,
  saveImage = FALSE,
  dpi = 300
)
```

Arguments

`identifier`	A single identifier for the query, either numeric or character.
`namespace`	A character string specifying the namespace for the request. Default is 'cid'.
`domain`	A character string specifying the domain for the request. Default is 'compound'.
`operation`	An optional character string specifying the operation for the request.
`output`	A character string specifying the output format. Possible values are 'JSON', 'JSONP', 'XML', 'CSV', 'TXT', and 'PNG'. Default is 'JSON'.
`searchtype`	An optional character string specifying the search type.
`property`	An optional character string specifying the property for the request.
`options`	A list of additional options for the request.
`saveFile`	A logical value indicating whether to save the output as a file. Default is FALSE.
`saveImage`	A logical value indicating whether to save the output as an image. Default is FALSE.
`dpi`	An integer specifying the DPI for image output. Default is 300.

Value

Depending on the output format, this function returns different types of content: JSON or JSONP format returns parsed JSON content. XML format returns an XML object. CSV format returns a data frame. TXT format returns a table. PNG format returns an image object or saves an image file.

Examples

```
get_pug_rest(identifier = "2244", namespace = "cid", domain = "compound", output = "JSON")
```

---

get_pug_view *Retrieve PUG View Data from PubChem*

Description

This function sends a request to the PubChem PUG View API to retrieve various types of data for a given identifier. It supports fetching annotations, QR codes, and more, with options for different output formats including JSON and SVG.

Usage

```
get_pug_view(
  annotation = NULL,
  identifier = NULL,
  domain = "compound",
  output = "JSON",
  heading = NULL,
  headingType = NULL,
  page = NULL,
  qrSize = "short",
  save = FALSE
)
```

Arguments

annotation	A character string specifying the type of annotation to retrieve.
identifier	A single identifier for the query, either numeric or character.
domain	A character string specifying the domain for the request. Default is 'compound'.
output	A character string specifying the output format. Possible values are 'JSON', 'XML', and 'SVG'. Default is 'JSON'.
heading	An optional character string for specifying a heading in the request.
headingType	An optional character string for specifying a heading type in the request.
page	An optional character string for specifying a page number in the request.
qrSize	A character string specifying the size of the QR code. Possible values are 'short' and 'long'. Default is 'short'.
save	A logical value indicating whether to save the output. Default is FALSE.

Value

Depending on the output format, this function returns different types of content: JSON or JSONP format returns parsed JSON content. SVG format returns an image object. For QR codes, it returns an image object or saves a PNG file.

Examples

```
get_pug_view(identifier = "2244", annotation = "linkout", domain = "compound")
```

`get_sdf` *Retrieve SDF Data from PubChem and Save as File*

Description

This function sends a request to PubChem to retrieve data in SDF format based on the specified parameters. It then saves the retrieved data as an SDF file in the current working directory.

Usage

```
get_sdf(
  identifier,
  namespace = "cid",
  domain = "compound",
  operation = NULL,
  searchtype = NULL,
  path = NULL,
  file_name = NULL,
  options = NULL
)
```

Arguments

`identifier`	A character or numeric value specifying the identifier for the request.
`namespace`	A character string specifying the namespace for the request. Default is 'cid'.
`domain`	A character string specifying the domain for the request. Default is 'compound'.
`operation`	An optional character string specifying the operation for the request.
`searchtype`	An optional character string specifying the search type.
`path`	A string indicating the path to the folder where the SDF files will be saved. Default is NULL (i.e., saves the file into a temporary folder).
`file_name`	A string. File name for downloaded SDF file. If NULL, "file" is used as the file name. Default is NULL.
`options`	Additional parameters to be passed to the `request`.

Value

NULL. The function saves the retrieved data as an SDF file in the current working directory and prints a message indicating the file's location.

Examples

```
get_sdf(
  identifier = "aspirin",
  namespace = "name",
  path = NULL
)
```

`get_sids` *Retrieve Substance IDs (SIDs) from PubChem*

Description

This function sends a request to PubChem to retrieve Substance IDs (SIDs) for a given identifier. It returns a tibble (data frame) with the provided identifier and the corresponding SIDs.

Usage

```
get_sids(
  identifier,
  namespace = "cid",
  domain = "compound",
  searchtype = NULL,
  options = NULL
)
```

Arguments

`identifier`	A numeric or character vector specifying the identifiers for the request.
`namespace`	A character string specifying the namespace for the request. Default is 'cid'.
`domain`	A character string specifying the domain for the request. Default is 'compound'.
`searchtype`	A character string specifying the search type. Default is NULL.
`options`	Additional arguments passed to `get_json`.

Value

A tibble (data frame) where each row corresponds to a provided identifier and its SID. The tibble has columns 'CID' and 'SID'.

Examples

```
get_sids(
  identifier = "aspirin",
  namespace = "name"
)
```

`get_substances` *Retrieve Substances from PubChem*

Description

This function sends a request to PubChem to retrieve substance data based on the specified parameters. It returns a list of substances corresponding to the provided identifiers.

Usage

```
get_substances(
  identifier,
  namespace = "sid",
  operation = NULL,
  searchtype = NULL,
  options = NULL
)
```

Arguments

`identifier`	A character or numeric vector specifying the identifiers for the request.
`namespace`	A character string specifying the namespace for the request. Default is 'sid'.
`operation`	Specifies the operation to be performed on the input records. For the 'compound' domain, possible operations include 'record', 'property', 'synonyms', 'sids', 'cids', 'aids', 'assaysummary', 'classification', 'xrefs', and 'description'. The available operations are domain-specific.
`searchtype`	Specifies the type of search to be performed. For structure searches, possible values are combinations of 'substructure', 'superstructure', 'similarity', 'identity' with 'smiles', 'inchi', 'sdf', 'cid'. For fast searches, possible values are combinations of 'fastidentity', 'fastsimilarity_2d', 'fastsimilarity_3d', 'fastsubstructure', 'fastsuperstructure' with 'smiles', 'smarts', 'inchi', 'sdf', 'cid', or 'fastformula'.
`options`	Additional parameters passed to `get_json`.

Value

A named list where each element corresponds to a substance retrieved from PubChem. The names of the list elements are based on the provided identifiers. If no substance is found for a given identifier, the corresponding list element will contain the string "No substance".

`get_synonyms` *Retrieve Synonyms from PubChem*

Description

This function sends a request to PubChem to retrieve synonyms for a given identifier. It returns a list of synonyms corresponding to the provided identifier.

Usage

```
get_synonyms(
  identifier,
  namespace = "cid",
  domain = "compound",
  searchtype = NULL,
  options = NULL
)
```

Arguments

`identifier`	A character or numeric value specifying the identifier for the request.
`namespace`	A character string specifying the namespace for the request. Default is 'cid'.
`domain`	A character string specifying the domain for the request. Default is 'compound'.
`searchtype`	A character string specifying the search type. Default is NULL.
`options`	Additional arguments passed to `get_json`.

Value

A list where each element corresponds to the synonyms retrieved from PubChem for the provided identifier. The names of the list elements are based on the provided identifier.

Examples

```
get_synonyms(
  identifier = "aspirin",
  namespace = "name"
)
```

ProjectCategory *Project Categories*

Description

List of project categories.

Usage

```
ProjectCategory
```

Format

An object of class `list` of length 11.

property_map *Property Map*

Description

Map of properties to their respective names.

Usage

```
property_map
```

Format

An object of class `list` of length 40.

pubchem_summary *Summarize Data from PubChem Based on Identifier*

Description

This function provides a comprehensive summary of data from the PubChem database for a given identifier. It can retrieve information about compounds, substances, assays, and additional properties, including synonyms and SDF files.

Usage

```
pubchem_summary(
  identifier,
  namespace = "cid",
  type = c("compound", "substance", "assay"),
  properties = NULL,
  include_synonyms = FALSE,
  include_sdf = FALSE,
  sdf_path = NULL,
  sdf_file_name = NULL,
  options = NULL
)
```

Arguments

identifier	A character string or numeric value representing the identifier for which the summary is required. It can be a compound ID (CID), substance ID (SID), assay ID (AID), or a name.
namespace	A character string specifying the namespace of the identifier. Possible values include 'cid' for compound ID, 'sid' for substance ID, 'aid' for assay ID, and 'name' for common names or synonyms.
type	A character vector indicating the type of data to retrieve. Possible values are "compound", "substance", and "assay". This parameter determines the kind of information fetched from PubChem.
properties	An optional vector of property names to retrieve for the given identifier. If specified, the function fetches these properties from PubChem.
include_synonyms	Logical; if TRUE, the function also retrieves synonyms for the given identifier.
include_sdf	Logical; if TRUE, the function downloads the Structure-Data File (SDF) for the given identifier.
sdf_path	An optional file path for saving the downloaded SDF file. If NULL and 'include_sdf' is TRUE, the file is saved into a temporary folder with the identifier as its name.
sdf_file_name	a character indicating the name of SDF file withoud ".sdf" extension. If NULL, default name is retrieved from `identifier` argument.
options	Additional arguments passed to internal functions.

Value

A list containing the requested data. The structure of the list depends on the parameters provided. It may include compound data, substance data, assay data, CIDs, SIDs, AIDs, synonyms, properties, and an SDF file path.

Examples

```
summary_data <- pubchem_summary(
  identifier = "aspirin",
  namespace = 'name',
  type = c("compound", "substance", "assay"),
  properties = "IsomericSMILES",
  include_synonyms = TRUE,
  include_sdf = TRUE
)
```

`request` *Request Function for PubChem API*

Description

Constructs a URL for the PubChem API based on the provided parameters.

Usage

```
request(
  identifier = NULL,
  namespace = "cid",
  domain = "compound",
  operation = NULL,
  output = "JSON",
  searchtype = NULL,
  options = NULL
)
```

Arguments

`identifier`	The identifier for the compound.
`namespace`	The namespace for the identifier (default: 'cid').
`domain`	The domain for the request (default: 'compound').
`operation`	The operation to be performed (default: NULL).
`output`	The desired output format (default: 'JSON').
`searchtype`	The type of search to be performed (default: NULL).
`options`	Additional parameters. Currently has no effect on the results.

Value

A constructed URL for the PubChem API.

Examples

```
request(
  identifier = "aspirin",
  namespace = "name"
)
```

Index