

Package: biorecap (via r-universe)

August 26, 2024

Title Retrieve and summarize bioRxiv preprints with a local LLM using ollama

Version 0.1.1

Description Retrieve and summarize bioRxiv preprints with a local LLM using ollama.

License MIT + file LICENSE

URL <https://stephenturner.github.io/biorecap/>

BugReports <https://github.com/stephenturner/biorecap/issues>

Depends R (>= 4.2.0)

Imports dplyr, ollamar (>= 1.2.1), rlang, rmarkdown, tidyRSS, tinytable

Suggests knitr, markdown, testthat (>= 3.0.0)

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Config/testthat/edition 3

Repository <https://stephenturner.r-universe.dev>

RemoteUrl <https://github.com/stephenturner/biorecap>

RemoteRef HEAD

RemoteSha 64d927a0db4cad746812c483eda7bc3edaf0c899

Contents

add_prompt	2
add_prompt_subject	3
add_summary	3
biorecap_report	4
build_prompt_preprint	5

build_prompt_subject	6
example_preprints	7
get_preprints	7
subjects	8
tt_preprints	8

Index	10
--------------	-----------

add_prompt	<i>Add prompt to a data frame of preprints</i>
------------	--

Description

Add prompt to a data frame of preprints

Usage

```
add_prompt(preprints, ...)
```

Arguments

preprints	Result from get_preprints() .
...	Additional arguments to build_prompt_preprint() .

Value

A data frame of bioRxiv preprints with a prompt added.

See Also

[build_prompt_preprint\(\)](#)

Examples

```
preprints <- get_preprints(subject=c("bioinformatics", "genomics"))
preprints <- add_prompt(preprints)
preprints
```

add_prompt_subject *Add prompts for an entire subject*

Description

Add prompts for an entire subject

Usage

```
add_prompt_subject(preprints, ...)
```

Arguments

preprints Output from [get_preprints\(\)](#) followed by [add_prompt\(\)](#) followed by [add_summary\(\)](#).
... Additional arguments to [build_prompt_subject\(\)](#).

Value

A tibble with a subject and prompt column.

Examples

```
subjects <-  
  example_preprints |>  
  dplyr::group_by(subject) |>  
  add_prompt_subject()  
subjects
```

add_summary *Generate a summary from a data frame of prompts*

Description

Generate a summary from a data frame of prompts

Usage

```
add_summary(preprints, model = "llama3.1")
```

Arguments

preprints Output from [get_preprints\(\)](#) followed by [add_prompt\(\)](#).
model A model available to Ollama (run `ollamar::list_models()`) to see what's available.

Value

A tibble, with a response column added.

Examples

```
## Not run:
# Individual papers
preprints <-
  get_preprints(c("genomics", "bioinformatics")) |>
  add_prompt() |>
  add_summary()
preprints

## End(Not run)
```

 biorecap_report

Create a report from bioRxiv preprints

Description

Create a report from bioRxiv preprints

Usage

```
biorecap_report(
  output_dir = ".",
  subject = NULL,
  nsentences = 2L,
  model = "llama3.1",
  use_example_preprints = FALSE,
  ...
)
```

Arguments

output_dir	Directory to save the report.
subject	Character vector of subjects to include in the report.
nsentences	Number of sentences to summarize each paper in.
model	The model to use for generating summaries. See ollamar::list_models() .
use_example_preprints	Use the example preprints data included with the package instead of fetching new data from bioRxiv. For diagnostic/testing purposes only.
...	Other arguments passed to rmarkdown::render() .

Value

Nothing; called for its side effects to produce a report.

Examples

```
## Not run:
output_dir <- tempdir()
biorecap_report(use_example_preprints=TRUE, output_dir=output_dir)
biorecap_report(subject=c("bioinformatics", "genomics", "synthetic_biology"),
                output_dir=output_dir)

## End(Not run)
```

build_prompt_preprint *Construct a prompt to summarize a paper*

Description

Construct a prompt to summarize a paper

Usage

```
build_prompt_preprint(
  title,
  abstract,
  nsentences = 2L,
  instructions = c("I am giving you a paper's title and abstract.",
                  "Summarize the paper in as many sentences as I instruct.",
                  "Do not include any preamble text to the summary",
                  "just give me the summary with no preface or intro sentence.")
)
```

Arguments

title	The title of the paper.
abstract	The abstract of the paper.
nsentences	The number of sentences to summarize the paper in.
instructions	Instructions to the prompt. This can be a character vector that gets collapsed into a single string.

Value

A string containing the prompt.

Examples

```
build_prompt_preprint(title="A great paper", abstract="This is the abstract.")
```

build_prompt_subject *Construct a prompt to summarize a set of papers from a subject*

Description

Construct a prompt to summarize a set of papers from a subject

Usage

```
build_prompt_subject(  
  subject,  
  title,  
  summary,  
  nsentences = 5L,  
  instructions =  
    c("I am giving you information about preprints published in bioRxiv recently.",  
      "I'll give you the subject, preprint titles, and short summary of each paper.",  
      "Please provide a general summary new advances in this subject/field in general.",  
      "Provide this summary of the field in as many sentences as I instruct.",  
      "Do not include any preamble text to the summary",  
      "just give me the summary with no preface or intro sentence.")  
)
```

Arguments

subject	The name of the subject.
title	A character vector of titles in the subject
summary	A character vector of the summaries of the paper provided by get_preprints() followed by add_prompt() followed by add_summary() .
nsentences	The number of sentences to summarize the subject in.
instructions	Instructions to the prompt. This can be a character vector that gets collapsed into a single string.

Value

A string containing the prompt.

Examples

```
title <- example_preprints |> dplyr::filter(subject=="bioinformatics") |> dplyr::pull(title)  
summary <- example_preprints |> dplyr::filter(subject=="bioinformatics") |> dplyr::pull(summary)  
build_prompt_subject(subject="bioinformatics", title=title, summary=summary)
```

example_preprints	<i>Example preprints with summaries</i>
-------------------	---

Description

Example preprints with summaries from August 6, 2024.

Usage

```
example_preprints
```

Format

A tibble returned from `get_preprints()` followed by `add_prompt()` followed by `add_summary()`.

Examples

```
example_preprints
```

get_preprints	<i>Get bioRxiv preprints</i>
---------------	------------------------------

Description

Get bioRxiv preprints

Usage

```
get_preprints(  
  subject = "all",  
  baseurl = "https://connect.biorxiv.org/biorxiv_xml.php?subject=",  
  clean = TRUE  
)
```

Arguments

subject	A character vector of valid biorxiv subjects. See subjects .
baseurl	The base URL for the biorxiv RSS feed. Default is <code>https://connect.biorxiv.org/biorxiv_xml.php?</code> . Do not change unless you know what you are doing.
clean	Logical; try to strip out graphical abstract information? If TRUE, this strips away any text between O_FIG and C_FIG, and the words graphical abstract from the abstract text in the RSS feed.

Value

A data frame of bioRxiv preprints.

Examples

```
preprints <- get_preprints(subject=c("bioinformatics", "genomics"))
preprints
```

subjects	<i>bioRxiv subjects</i>
----------	-------------------------

Description

Names of subjects with RSS feeds in biorXiv

Usage

```
subjects
```

Format

A character vector

References

<https://www.biorxiv.org/alertsrss>

Examples

```
subjects
```

tt_preprints	<i>Create a markdown table from prepreprint summaries</i>
--------------	---

Description

Create a markdown table from prepreprint summaries

Usage

```
tt_preprints(preprints, cols = c("title", "summary"), width = c(1, 3))
```


Arguments

<code>preprints</code>	Output from <code>get_preprints()</code> followed by <code>add_prompt()</code> followed by <code>add_summary()</code> .
<code>cols</code>	Columns to display in the resulting markdown table.
<code>width</code>	Vector of relative widths equal to <code>length(cols)</code> .

Value

A tinytable table.

Examples

```
# Use built-in example data
example_preprints
tt_preprints(example_preprints[1:2,])
```

Index

* datasets

example_preprints, 7
subjects, 8

add_prompt, 2
add_prompt(), 3, 6, 7, 9
add_prompt_subject, 3
add_summary, 3
add_summary(), 3, 6, 7, 9

biorecap_report, 4
build_prompt_preprint, 5
build_prompt_preprint(), 2
build_prompt_subject, 6
build_prompt_subject(), 3

example_preprints, 7

get_preprints, 7
get_preprints(), 2, 3, 6, 7, 9

ollamar::list_models(), 4

rmarkdown::render(), 4

subjects, 7, 8

tt_preprints, 8