





- A Collection of News Media Scrapers in R

For UseRs

```
pb_deliver(  
  "www.test.com"  
)
```

generates

	url	expanded_url	domain	status	datetime	author	headline	text	misc
1	https://www.test1.com	https://www.te...	test1.com	200	2026-07-07 11:45:27	Author name	Article headline	Article text	<data.frame[1 x 0]>





- A Collection of News Media Scrapers in R

For UseRs

```
pb_deliver(  
  "www.test.com"  
)
```

generates

	url	expanded_url	domain	status	datetime	author	headline	text	misc
1	https://www.test1.com	https://www.te...	test1.com	200	2026-07-07 11:45:27	Author name	Article headline	Article text	<data.frame[1 x 0]>

```
pb_collect(  
  "www.test.com"  
)
```

generates

	url	expanded_url	domain	status	content_raw
1	https://www.test1.com	https://www.te...	test1.com	200	<!DOCTYPE html> <html lang="en"...





- A Collection of News Media Scrapers in R

For UseRs

```
pb_deliver(  
  pb_collect(  
    c("www.test1.com",  
      "www.test2.com",  
      ...)  
  )  
)
```

generates

	url	expanded_url	domain	status	datetime	author	headline	text	misc
1	https://www.test1.com	https://www.te...	test1.com	200	2026-07-07 11:45:27	Author name	Article headline	Article text	<data.frame[1 x 0]>
2	https://www.test2.com	https://www.te...	test2.com	200	2026-07-07 11:45:27	Author name	Article headline	Article text	<data.frame[1 x 0]>
3	...	...	...	...	...	...	...	...	...





- A Collection of News Media Scrapers in R

For Users

```
pb_deliver(  
  pb_collect(  
    c("www.test1.com",  
      "www.test2.com",  
      ...)  
  )  
)
```

For Developers

```
pb_find_rss("test.com")  
test_df = pb_collect("test.com/feed")  
use_new_parser(  
  x = test_df$expanded_url,  
  rss = "www.test.com/feed",  
  test_data = test_df  
)
```





- A Collection of News Media Scrapers in R

For Users

```
pb_deliver(  
  pb_collect(  
    c("www.test1.com",  
      "www.test2.com",  
      ...)  
  )  
)
```

For Developers

```
pb_find_rss("test.com")  
test_df = pb_collect("test.com/feed")  
use_new_parser(  
  x = test_df$expanded_url,  
  rss = "www.test.com/feed",  
  test_data = test_df  
)
```

generates

```
#' @export  
pb_deliver_paper.test_com ← function(x,  
                                     verbose= NULL,  
                                     pb, ... ) {  
  ...  
  raw_html ← rvest::read_html(x$content_raw)  
  datetime ← html %>% rvest::html_element(...  
  headline ←html %>% rvest::html_element(...  
  author ←html %>% rvest::html_element(...  
  text ←html %>% rvest::html_elements(...  
  ...  
}
```





- A Collection of News Media Scrapers in R

For UseRs

```
pb_deliver(  
  pb_collect(  
    c("www.test1.com",  
      "www.test2.com",  
      ...)  
  )  
)
```

For Developers

```
pb_find_rss("test.com")  
test_df = pb_collect("test.com/feed")  
use_new_parser(  
  x = test_df$expanded_url,  
  rss = "www.test.com/feed",  
  test_data = test_df  
)
```

generates

```
#' @export  
pb_deliver_paper.test_com ← function(x,  
                                     verbose= NULL,  
                                     pb, ... ) {  
  
  ...  
  raw_html ← rvest::read_html(x$content_raw)  
  
  datetime ← html %>% rvest::html_element(...  
  headline ←html %>% rvest::html_element(...  
  author ←html %>% rvest::html_element(...  
  text ←html %>% rvest::html_elements(...  
  ...  
}
```

LLM assisted
workflow

```
> /new-parser test.com  
>> /fix-parser test.com
```





- A Collection of News Media Scrapers in R

For UseRs

```
pb_deliver(  
  pb_collect(  
    c("www.test1.com",  
      "www.test2.com",  
      ...)  
  )  
)
```

For Developers

```
pb_find_rss("test.com")  
test_df = pb_collect("test.com/feed")  
use_new_parser(  
  x = test_df$expanded_url,  
  rss = "www.test.com/feed",  
  test_data = test_df  
)
```

generates

```
#' @export  
pb_deliver_paper.test_com ← function(x,  
                                     verbose= NULL,  
                                     pb, ... ) {  
  
  ...  
  raw_html ← rvest::read_html(x$content_raw)  
  
  datetime ← html %>% rvest::html_element(...  
  headline ←html %>% rvest::html_element(...  
  author ←html %>% rvest::html_element(...  
  text ←html %>% rvest::html_elements(...  
  ...  
}
```

LLM assisted
workflow

```
> /new-parser test.com  
  
> /fix-parser test.com
```

generates

fixes

