Exercises — Regex Hacker

The way is lit. The path is clear.
We require only the strength to follow it.

Copyright

This document is for internal use at EPITA ([website](#)) only.

Copyright © 2022-2023 Assistants `<assistants@tickets.assistants.epita.fr>`

The use of this document must abide by the following rules:
- ▷ You downloaded it from the assistants' intranet.*
- ▷ This document is strictly personal and must **not** be passed onto someone else.
- ▷ Non-compliance with these rules can lead to severe sanctions.

Contents

File Tree

```
hacker_news/
├── input.html
├── news.sed   (to submit)
└── output.txt
```

Reminder : Grant execution permission to your scripts before pushing them

1 Goal

This exercise's goal is to output the main page of the famous website Hacker News. You will have to output each title, surrounded by ** and followed by its url one line below, itself followed by an empty newline like so:

```
**Title 1**
url

**Title 2**
url
```

All 30 article titles should be printed.

You have to submit one single `sed` script called `news.sed`! You can assume that it will be used in the same way as the example below.

You can easily fetch the html page with commands like `curl` or `wget` to test, check their man page for more information.

2 Example

You will obviously have different titles depending on the time, this is a sample output.

```
42sh$ curl -s https://news.ycombinator.com/ | sed -n -f news.sed
**Cybersecurity Incident Involving Consumer Information**
https://investor.equifax.com/news-and-events/news/2017/09-07-2017-213000628

**Beej's Guide to C Programming**
http://beej.us/guide/bgc/

**$257M: Filecoin Breaks Record for ICO Funding**
https://www.coindesk.com/257-million-filecoin-breaks-time-record-ico-funding/

[... And many more lines ...]
```

We also provide you some files that can be useful to test your script.

```
42sh$ cat input.html | sed -n -f news.sed | diff - output.txt
42sh$ echo $?
0
```

The way is lit. The path is clear. We require only the strength to follow it.