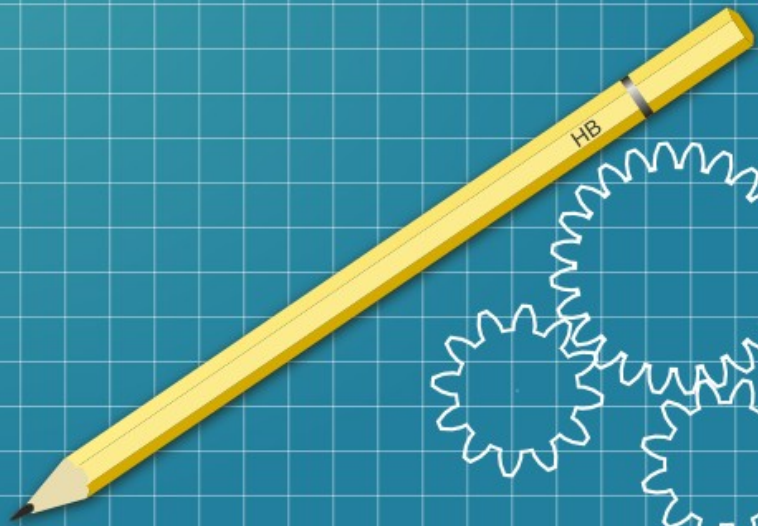
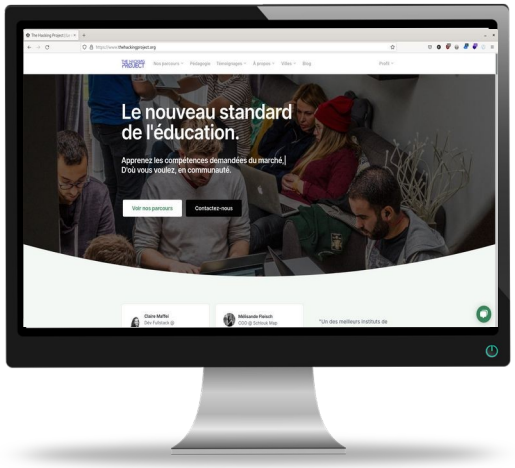
THP Search Engine



How this API works



POST https://thp-search.herokuapp.com/search
{ query : awesome+query }

```
JSON {
      "query":"awesome query",
      "results":[
        {
          "title":"most significant result's title",
          "url":"most significant result's url"
        },
        ..............
        {
          "title":"less significant result's title",
          "url":"less significant result's url"
        }
      ]
    }
```

How this Backend works

Based on Ruby on Rails 7

1. The « Scraper » class :

Based on Selenium, this class contains all the methods to

- log in to "The Hacking Project" platform
- retrieve all available course urls for the registered account
- retrieve course content separated into "title", "subheadings" and "contents"

2. The « Db » class :

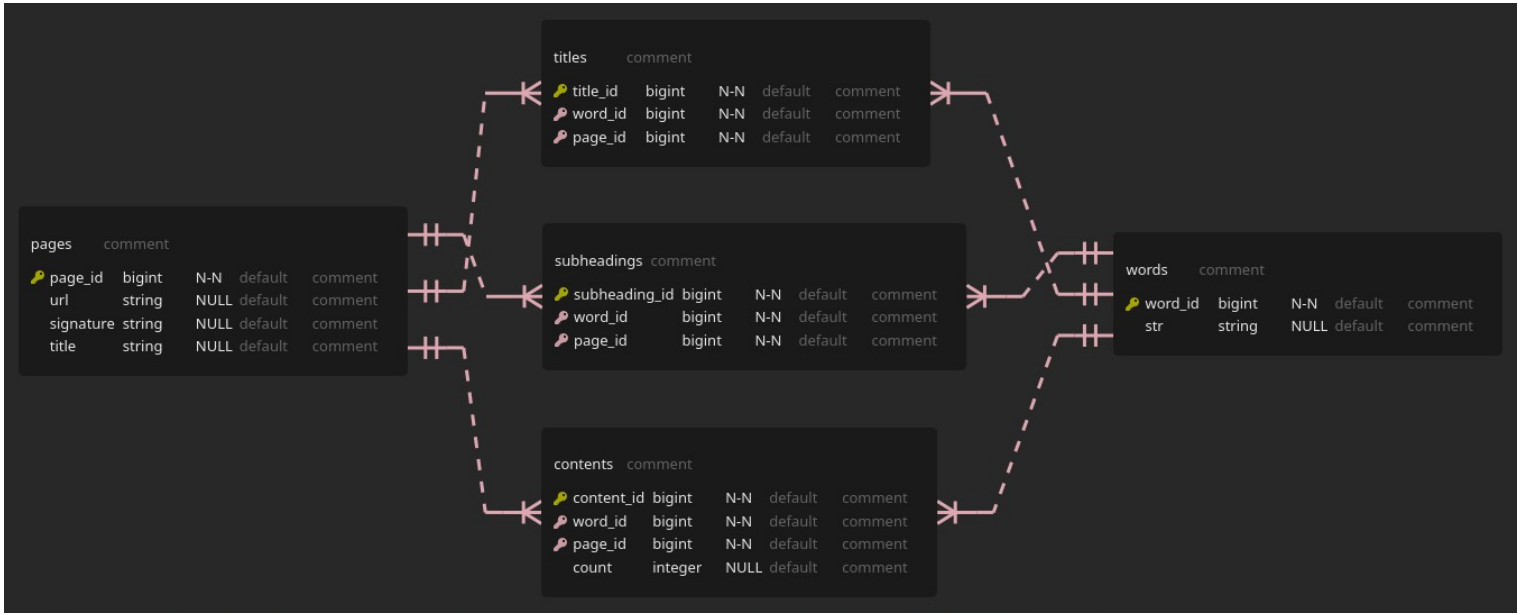
Automates the database population process.

3. The « Search » controller :

Process the POST request sent to /search and send the response in JSON format.

How this database works

Table pages :
Stores « url », « title » and the « signature » which is the digest of the article, this for all pages founded.

Table words :
Stores each word founded in articles. Each word is uniq.

Tables titles, subheadings and contents :
Are join tables for the Many-to-Many relationship between « pages » and « words ».