

Un scraper d'emails

Le scraping est une technique qui consiste à extraire des données spécifiques depuis des sites web lambda. Notre scraper d'emails vous servira à récolter toutes les adresses emails apparaissant dans le code source d'une page web donnée. Le code du scraper vous est fourni (vous devrez toutefois bien le comprendre), votre travail se situera essentiellement dans ce que nous ferons des emails trouvés.

Travail à faire :

1. Pour bien comprendre le fonctionnement du scraper d'emails qui vous est fourni, commencez par écrire un fichier **test_scraper.php** contenant le code suivant :

```
<?php
    include 'email_scraper.php';
    $url = 'https://github.com/nyxgeek/username-lists/blob/master/usernames-top100/usernames_gmail.com.txt';
    $emails = scrape_email($url);
    echo implode($emails, '<br>');
?>
```

Chargez ce fichier dans votre navigateur, puis lisez bien le code du fichier **email_scraper.php** avant de passer à la suite.

2. Vous allez ensuite créer une base de données composée d'une table **Emails** contenant un unique champ « *email* » de type varchar(128) et qui sera aussi la **clé primaire** de la table.
3. A l'aide du *framework* **Medoo** (dont le fichier vous est fourni) vous allez créer une page web **myscraper.php** dont la fonction sera d'ajouter en base de données toutes adresses emails trouvées à une url donnée. *Si vous n'avez jamais utilisé Medoo jusqu'à présent, consultez la dernière section de cet exercice.*
4. Ajoutez désormais un formulaire à votre page web dans lequel vous pourrez saisir une url dans un input. En soumettant ce formulaire, votre *scraper* ira extraire les adresses emails trouvées à l'url donnée et les ajoute en base de données. Une fois le travail terminé, il recharge le formulaire.
5. Enfin, après chaque *scraping*, vous afficherez le nombre d'adresses emails trouvées ainsi que la liste de ces emails dans votre page dans un joli tableau, sous le formulaire.

6. Pour terminer proprement l'exercice, passez un petit coup de CSS sur votre code afin de le rendre plus présentable.

Bonus :

- Essayez d'optimiser au mieux le requêtage en essayant d'envoyer un maximum d'informations en un minimum de requêtes.

Medoo :

Medoo est un *framework* PHP ultra léger orienté objets vous permettant d'écrire facilement vos composants d'accès à la base de données de manière sécurisée, en vous protégeant de différentes formes d'injections.

Si vous n'avez pas fait l'exercice de la gestion d'une collection de cartes Magic, ce n'est pas grave, vous apprendrez à utiliser **Medoo** maintenant. Rendez-vous sur le site du *framework* **Medoo** pour comprendre comment l'utiliser. <https://medoo.in/>

Commencez par suivre les instructions dans le « *Get started* » (pas besoin de télécharger **Medoo**, le fichier **Medoo.php** vous est déjà fourni).

La documentation est disponible à cette adresse : <https://medoo.in/doc>

Si vous avez besoin d'exemples d'utilisation de **Medoo**, reportez-vous au boss **Gestion d'une collection de cartes Magic**. Vous trouverez tout ce dont vous avez besoin dans le code.