

# **Analiza paginii web pentru extragerea adreselor de email: O abordare computațională**

Autor: Șovago Rareș-Andrei  
Coordonator: conf univ dr Sabo Cosmin  
Universitatea Tehnică din Cluj-Napoca  
Facultate de Științe  
Specializarea: Informatică

# Obiectivul proiectului:

Extragerea adreselor de email de pe site-urile web vizitate pentru o analiză ulterioară și o stocare agreata din punct de vedere legislativ (IGPR), oferind o serie de funcționalități menite să ajute utilizatorul final.



# Motivatia proiectului:

Componentă esențială pentru colectarea informațiilor referitoare la produse, opinii, link-uri de marketing și concepte de producție/dezvoltare.

Aceste date reprezintă elemente vitale pentru consolidarea competitivității în cadrul sectorului privat.



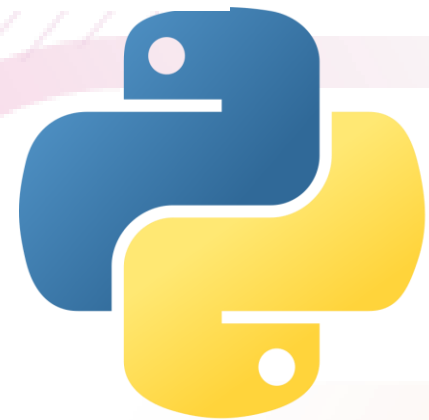
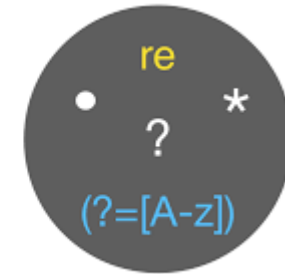
# Tehnologii utilizate



# Playwright



Python  
regex  
Regular Expressions



# Playwright

Navigare WEB

Depth

Extragerea elementelor specifice



```
from playwright.sync_api import sync_playwright
```

# HeidiSQL

Clienți+formular opinie  
phpMyAdmin



# Regex (expresii regulate)

```
page_content = page.query_selector("body")
```

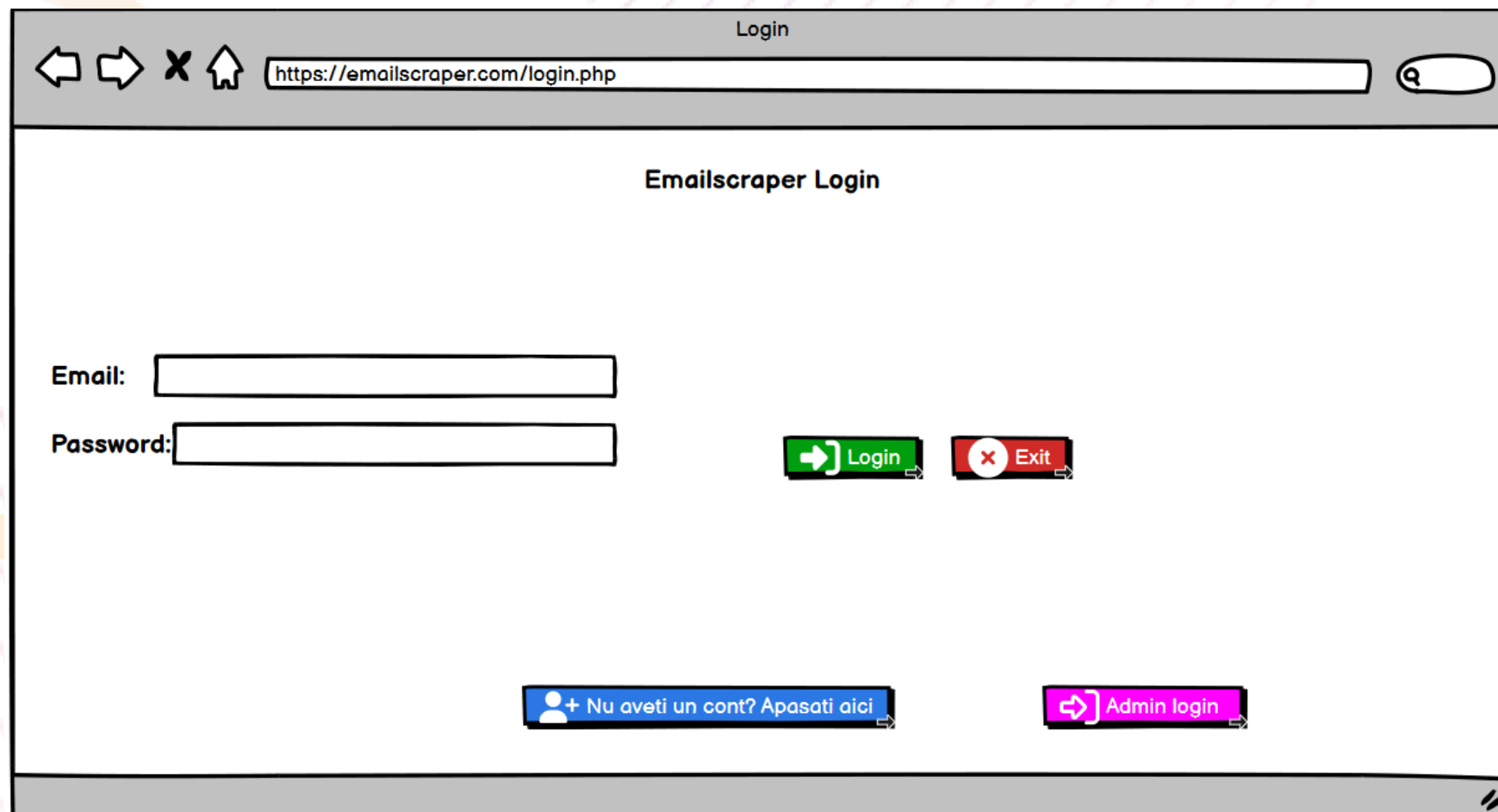
```
email_regex = re.compile(r"[a-zA-Z0-9_+-.]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-.]+")
```

```
urls = re.findall(r'href=["\'](http[s]?://[^\s"\<>]+)', cont)
```

# Accesul la aplicație

Client logat

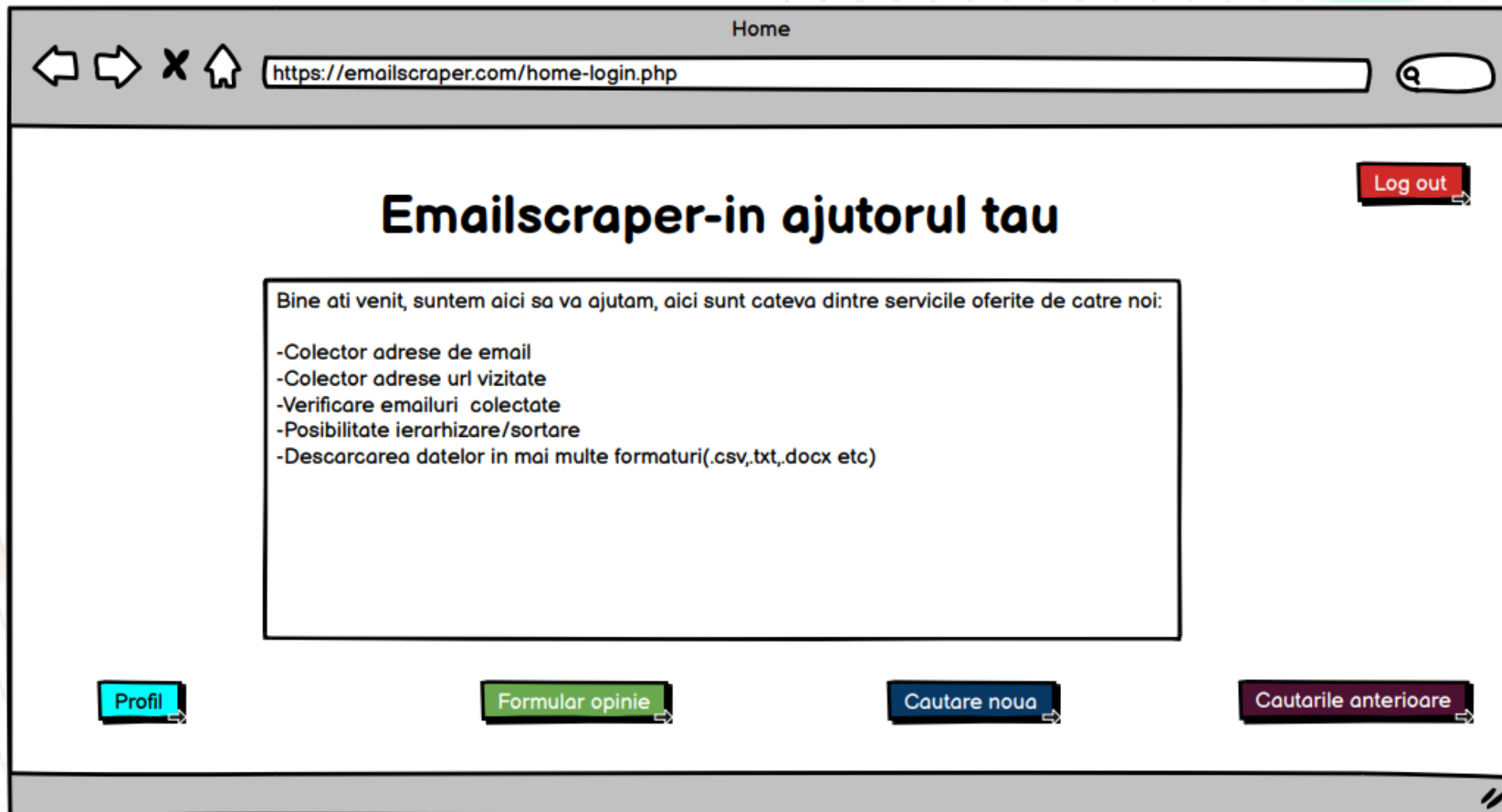
Administrator



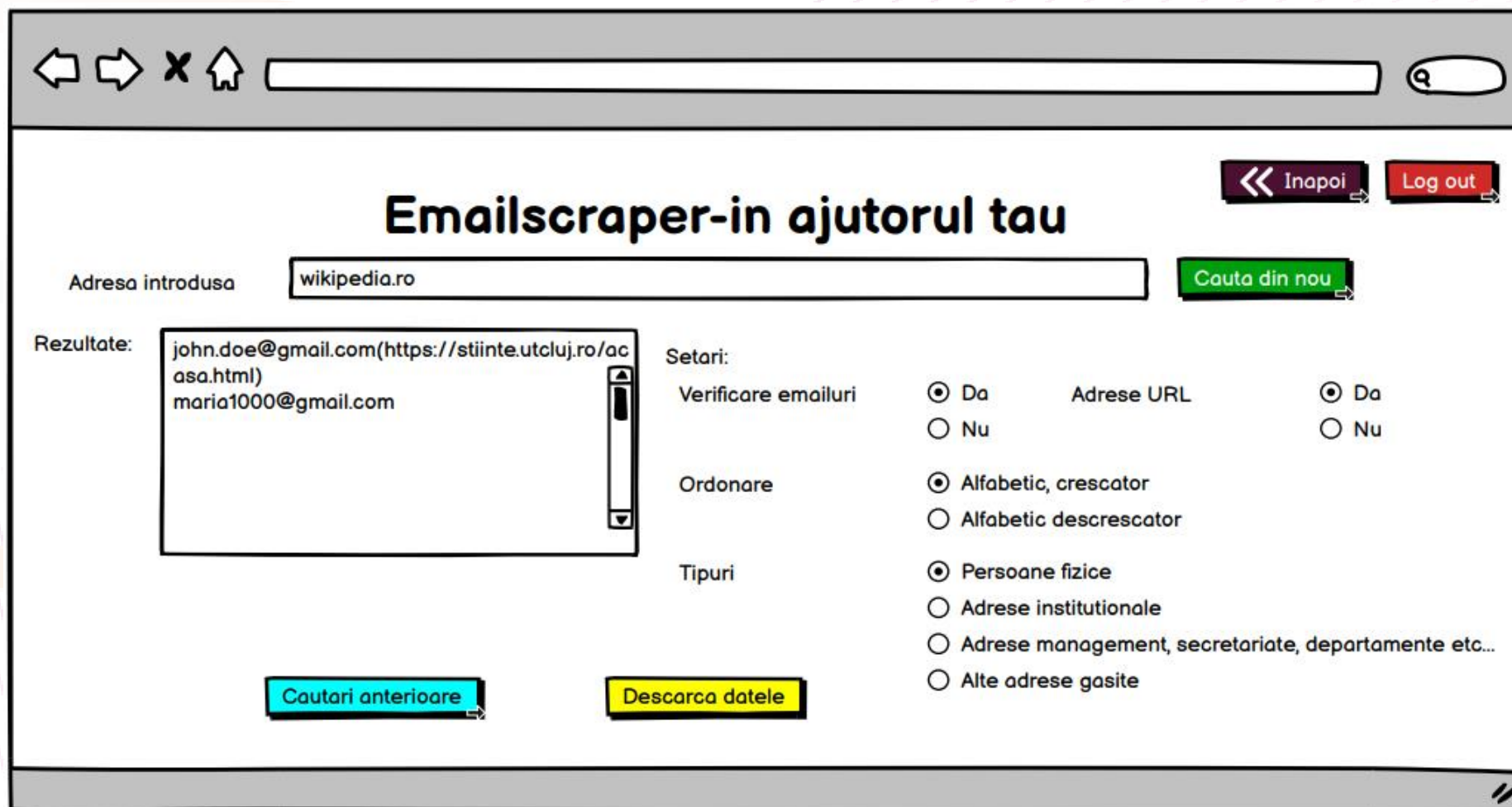
The screenshot shows a web browser window with the title "Login" and the address bar containing "https://emailscraper.com/login.php". The page content is titled "Emailscraper Login" and features two input fields: "Email:" and "Password:". Below the input fields are four buttons: a green "Login" button with a right-pointing arrow, a red "Exit" button with a white 'X', a blue button with a person icon and the text "+ Nu aveti un cont? Apasati aici", and a pink button with a right-pointing arrow and the text "Admin login".



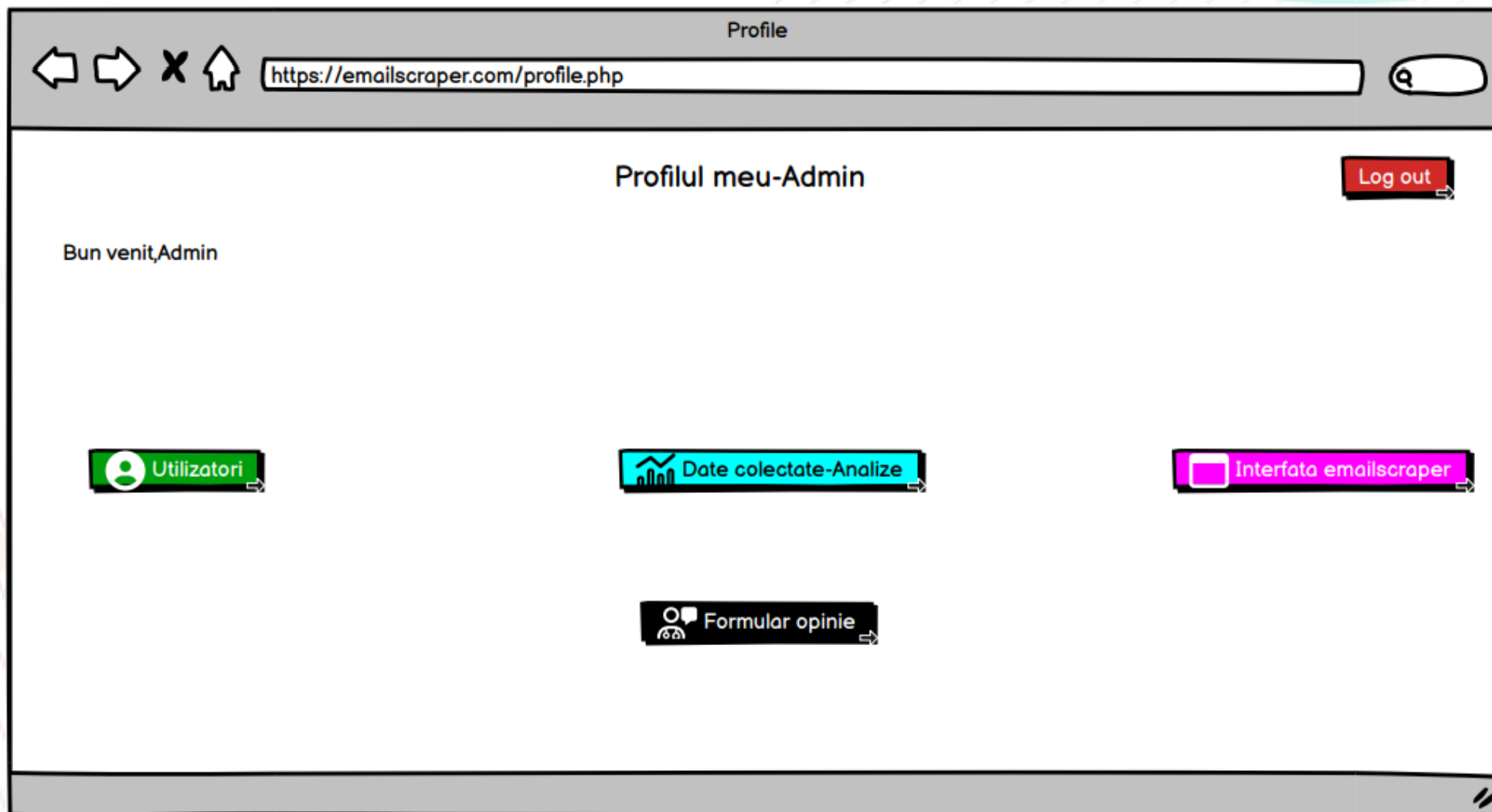
# Meniul utilizatorului



# Setările aplicației



# Interfața administratorului



# Idei de îmbunătățire

Eficiență + Eficacitate

Extragerea de tipuri de date multiple

Îmbunătățirea sistemului de siguranță

Descărcarea datelor în mai multe forme (ex: docx, pdf, db3)



# Concluzie

“Mama e hotărâtă să se familiarizeze cu noua tehnologie. Mi-a scris un e-mail pe computer, l-a printat, l-a pus în plic și mi l-a trimis prin poștă.” John O'Farrell

