

# Web Scraping API Service

## Test tehnic

### Introducere

Această aplicație permite utilizatorilor să introducă o adresă URL și să selecteze elementele HTML pe care doresc să le extragă (titluri - h3, paragrafe - p, div-uri - div) de pe pagina web asociată URL-ului. De asemenea, utilizatorii au opțiunea de a analiza sentimentul textului extras. Aplicația utilizează Puppeteer pentru a efectua web scraping și Sentiment pentru analiza sentimentului textului.

### Endpoints API

#### 1. `/scrape`

Acest endpoint este folosit pentru a efectua web scraping și analiză a paginilor web.

```
{
  "url": "https://wsa-test.vercel.app/",
  "scrapeElements": ["h3", "p", "div"],
  "analyzeSentiment": true
}
```

- `url`: Adresa URL a paginii web de pe care se va efectua scraping.
- `scrapeElements`: Elementele HTML de pe care se dorește să se facă scraping (h3, p, div).
- `analyzeSentiment`: Dacă se dorește sau nu analiza sentimentului textului extras.

Răspunsul va fi un obiect JSON care conține datele extrase și, dacă este cazul, analiza sentimentului.

Exemplu de răspuns:

```
{
  "headings": ["Titlu 1", "Titlu 2", ...],
  "paragraphs": ["Paragraf 1", "Paragraf 2", ...],
  "divs": ["Div 1", "Div 2", ...],
  "sentiment": {
    "score": 2,
    "comparative": 0.1,
    "tokens": ["cuvânt1", "cuvânt2", ...],
```

```
"positive": [...],  
"negative": [...],  
"scorePerWord": 0.02  
},  
"wordCount": 100  
}
```

## 2. `GET /`

Acest endpoint servește fișierul front-end, pagina HTML principală.

## Rulare aplicație

1. Începe prin instalarea dependențelor cu comanda `npm install` în directorul proiectului.
2. Node.js și Puppeteer sunt necesare în rularea aplicației.
3. Pentru a rula aplicația, utilizează comanda `node api.js`.
4. Accesează aplicația într-un browser web la adresa `http://localhost:3000`.

## Caracteristici

- Aplicația utilizează Puppeteer pentru a efectua web scraping în timp real pe paginile web specificate de utilizatori.
- Opțiunea de analiză a sentimentului folosește biblioteca Sentiment pentru a evalua conținutul textului extras.
- Front-end-ul aplicației este simplu și intuitiv, permițând utilizatorilor să introducă URL-ul și să selecteze elementele dorite pentru scraping.

## Experiența de învățare

În timpul dezvoltării acestei aplicații, am învățat cum să utilizez Puppeteer pentru a efectua web scraping și cum să integrez analiza sentimentului în aplicație.

A fost o experiență valoroasă care mi-a permis să învăț mai multe despre tehnologii web și să dezvolt abilități de dezvoltare web full-stack. Această aplicație poate fi extinsă și îmbunătățită cu mai multe funcționalități în viitor.

## **Concluzie**

Această aplicație de web scraping și analiză reprezintă o soluție utilă pentru extragerea și analiza datelor de pe paginile web. Este un exemplu bun de utilizare a tehnologiilor web și poate fi extinsă pentru a satisface nevoile specifice ale utilizatorilor. Cu o documentație detaliată și un front-end prietenos, această aplicație poate fi folosită în diverse scopuri, inclusiv cercetare, analiză de piață și multe altele.