

소프트웨어 나눔 축제

AnA - Node.js로 크롤링해보기

크롤링

puppeteer

1. puppeteer
2. 프로젝트

puppeteer은 chrome 또는 chromium을 제어하는 API를 제공하는 라이브러리입니다.

API : 프로그램끼리의 연결망
라이브러리 : 타인이 만들어 둔 코드 묶음₁

크롤링

puppeteer

설치해봅시다

npm i puppeteer

1. puppeteer

2. 프로젝트

크롤링

puppeteer

사용해봅시다

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('https://sunrint.sen.hs.kr/');
  await page.screenshot({path: 'example.png'});

  await browser.close();
})();
```

1. puppeteer

2. 프로젝트

크롤링

puppeteer

모듈 가져오기

```
const puppeteer = require('puppeteer');
```

```
(async () => {  
  const browser = await puppeteer.launch();  
  const page = await browser.newPage();  
  await page.goto('https://sunrint.sen.hs.kr/');  
  await page.screenshot({path: 'example.png'});  
  
  await browser.close();  
})();
```

크롤링

puppeteer

런처실행

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('https://sunrint.sen.hs.kr/');
  await page.screenshot({path: 'example.png'});

  await browser.close();
})();
```

크롤링

puppeteer

새페이지 만들기

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('https://sunrint.sen.hs.kr/');
  await page.screenshot({path: 'example.png'});

  await browser.close();
})();
```

크롤링

puppeteer

해당 페이지 이동

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('https://sunrint.sen.hs.kr/');
  await page.screenshot({path: 'example.png'});

  await browser.close();
})();
```


크롤링

puppeteer

페이지 스크린샷 저장

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('https://sunrint.sen.hs.kr/');
  await page.screenshot({path: 'example.png'});

  await browser.close();
})();
```

크롤링

puppeteer

종료

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('https://sunrint.sen.hs.kr/');
  await page.screenshot({path: 'example.png'});

  await browser.close();
})();
```

크롤링

puppeteer

조금더 복잡한 예제입니다.

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch({ headless: false });
  try {
    const page = await browser.newPage();
    await page.goto('https://sunrint.sen.hs.kr/');
    const text = await page.evaluate(() => {
      const score = document.querySelector('.menu');
      if (score) {
        return score.textContent;
      }
    });
    console.log(text.trim());
  } catch (err) {
    console.log(err);
  }
  await browser.close();
})();
```

1. puppeteer

2. 프로젝트

크롤링

puppeteer

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch({ headless: false });
  try {
    const page = await browser.newPage();
    await page.goto('https://sunrint.sen.hs.kr/');
    const text = await page.evaluate(() => {
      const score = document.querySelector('.menu');
      if (score) {
        return score.textContent;
      }
    });
    console.log(text.trim());
  } catch (err) {
    console.log(err);
  }
  await browser.close();
})();
```

시각화

1. puppeteer

2. 프로젝트

크롤링

puppeteer

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch({ headless: false });
  try {
    const page = await browser.newPage();
    await page.goto('https://sunrint.sen.hs.kr/');
    const text = await page.evaluate(() => {
      const score = document.querySelector('.menu');
      if (score) {
        return score.textContent;
      }
    });
    console.log(text.trim());
  } catch (err) {
    console.log(err);
  }
  await browser.close();
})();
```

원하는 데이터 얻기

1. puppeteer

2. 프로젝트

크롤링

puppeteer

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch({ headless: false });
  try {
    const page = await browser.newPage();
    await page.goto('https://sunrint.sen.hs.kr/');
    const text = await page.evaluate(() => {
      const score = document.querySelector('.menu');
      if (score) {
        return score.textContent;
      }
    });
    console.log(text.trim());
  } catch (err) {
    console.log(err);
  }
  await browser.close();
})();
```

문서 객체 모델(DOM)

크롤링

문서 객체 모델(DOM)

```
<iframe id="backgroundImage" src=""></iframe>
```

```
document.querySelector('#backgroundImage');
```

1. puppeteer
2. 프로젝트

크롤링

프로젝트

1. puppeteer
2. 프로젝트

교재에 나와있는 프로젝트들을 진행해봅시다