

Week4 - Book2 : PDF파일 분할하기

20200191 김동연

Assignment 1

Chat GPT를 이용해서 자율적인 범위로
3등분 하기



Web Assembly를 활용한 웹 어플리케이션...
PDF



자, 이 pdf를 0~19페이지, 20~31페이지, 그리고 나머지 이렇게
3개의 PDF로 분할해줘



PDF가 성공적으로 세 개의 파일로 분할되었습니다:

1. [0~19 페이지](#)
2. [20~31 페이지](#)
3. [나머지 페이지](#)

필요하시면 링크를 통해 다운로드하실 수 있습니다. [\[~\]](#)

Assignment 1

Chat GPT를 이용해서 자율적인 범위로
3등분 하기



Web_Assembly_Split_Corrected_Part_1.p...
20/20페이지

Web_Assembly_Split_Final_Part_2.pdf
1/12페이지

WebAssembly_split_Final_Part_3.pdf
1/9페이지

Assignment 1

3등분한 PDF 파일 중 하나를 ChatGPT를
이용하여 이미지로 변환하기

→ 해상도가 굉장히 낮아진 걸 볼 수 있음

4.2 데이터 로드 및 렌더링 성능 테스트

2000개의 데이터를 로드하고 DOM Element를 생성하여 화면에 렌더링 시간을 측정하기 위해 숫자와 텍스트 형태의 9개 필드를 포함한 JSON 데이터를 준비했다.

이를 이용하여 JavaScript와 Web Assembly 버전의 성능을 비교하기 위해 해당 작업을 총 다섯 번씩 수행하여 평균 시간을 측정했다. JavaScript의 경우 평균 1,727ms Web Assembly의 경우 평균 1,475ms의 시간이 소요되었다.

이 결과 데이터 로드기능에서 Web Assembly는 JavaScript에 비해 대략 1.17배 빠른 성능을 보여주었다.

	A	B	C	D	E	F	G	H
1	Ctrl+F 보다 더 똑똑하게: 웹페이지 콘텐츠에 직접 링크하기	https://alfy.blog/2024/10/19/linking-directly-to-web-page-content.html						
2	GN+: Flutter를 Flock으로 포크합니다	https://flutterfoundation.dev/blog/posts/we-are-forking-flutter-this-is-why/						
3	글을 쓰는 사람과 그렇지 않은 사람	https://paulgraham.com/writes.html						
4	기술 업계의 거시적 트렌드 2024.10 by Thoughtworks	https://www.thoughtworks.com/insights/blog/technology-strategy/macro-trends-tech-industry-october-2024						
5	[번역] 소규모 언어 모델(SLM)에 대한 종합적인 연구	https://discuss.pytorch.kr/t/slm-small-language-models-survey-measurements-and-insights/5292?utm_source=geeknev						
6	GN+: Apple, M4 칩 탑재한 새로운 색상의 iMac 출시	https://www.apple.com/kr/newsroom/2024/10/apple-introduces-new-imac-supercharged-by-m4-and-apple-intelligence/						
7	Thoughtworks Technology Radar, Volume 31 공개	https://www.thoughtworks.com/radar						
8	Show GN: New Swagger-UI combined with Cloud TtypeScrit Editor and embedded SDK	https://news.hada.io/topic?id=17480						
9	GN+: 오픈 워싱 – 기업들이 오픈 소스인척하는 이유	https://www.theregister.com/2024/10/25/opinion_open_washing/						
10	DuckDB 사용법(DuckDB Python + Jupyter Lab)	https://zsza.github.io/data-engineering/2024/10/25/duckdb/						
11								
12								
13								

Assignment 2

원하는 웹사이트의 특정 파트를 웹스크래핑 해보기
(예: 국민대학교 소용대 공지사항 상위 10개 등).

Google Sheets 내의 App Script를 사용해서 진행
한다. 진행 과정에서 ChatGPT의 도움을 받는다.

**[https://docs.google.com/spreadsheets/
d/15yS4VLGhjNzZmHqA5axgeANrpxcdcLXFmsX8Dh3P9Jw/edit?
gid=0#gid=0](https://docs.google.com/spreadsheets/d/15yS4VLGhjNzZmHqA5axgeANrpxcdcLXFmsX8Dh3P9Jw/edit?gid=0#gid=0)**

Assignment 2

원하는 웹사이트의 특정 부분을 웹스크래핑 해보기
(예: 국민대학교 소용대 공지사항 상위 10개 등).
Google Sheets 내의 App Script를 사용해서 진행
한다. 진행 과정에서 ChatGPT의 도움을 받는다.

Apps Script 웹 스크래핑 실습 배포 ? ☰ 🌐

파일 ⌵ ⌵ ⌵ 실행 디버그 fetchDataFromGeekNews 실행 로그

Code.gs

라이브러리 +

서비스 +

```
1 function fetchDataFromGeekNews() {
2   const url = "https://news.hada.io/";
3   // Step 1
4   Logger.log("Fetching data from: " + url);
5
6   try {
7     const response = UrlFetchApp.fetch(url);
8     // Step 2
9     Logger.log("Response received");
10
11     const html = response.getContentText();
12     // HTML 확인
13     Logger.log(html);
14     // HTML을 파싱해 데이터 추출
15     const data = parseGeekNewsHTML(html);
16
17     // Step 3
18     Logger.log("Data parsed successfully");
19     Logger.log(data); // 추출된 데이터 확인
20
21     const sheet = SpreadsheetApp.getActiveSpreadsheet().getActiveSheet();
22     sheet.clear(); // 이전 데이터를 지우고 업데이트
23
24     // 데이터를 각 행에 저장
25     data.forEach((item, index) => {
26       sheet.getRange(index + 1, 1).setValue(item.title); // A열에 제목
27       sheet.getRange(index + 1, 2).setValue(item.link); // B열에 링크
28     });
29     // Step 4
30     Logger.log("Data written to the sheet");
31   } catch (error) {
32     Logger.log("Error occurred: " + error.message); // Error handling
33   }
34 }
```

Assignment 2

원하는 웹사이트의 특정 파트를 웹스크래핑 해보기
(예: 국민대학교 소용대 공지사항 상위 10개 등).
Google Sheets 내의 App Script를 사용해서 진행
한다. 진행 과정에서 ChatGPT의 도움을 받는다.

배포	함수	유형	시작 시간	기간	상태	
Head	fetchDataFromGeekNews	시간 기반	2024. 10. 28. 오후 9:56:57	1.901초	완료됨	<div><div></div><div></div></div>
Cloud 로그						
2024. 10. 28. 오후 9:56:58	정보	Fetching data from: https://news.hada.io/				
2024. 10. 28. 오후 9:56:58	정보	Response received				
2024. 10. 28. 오후 9:56:58	정보	Logging output too large. Truncating output. <!DOCTYPE html><html lang="ko"><head><meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1" /><script src="/vendor/markdown-it.min.js"></script><link rel="stylesheet" href="/vendor/highlight-default.min.css"><script src="/vendor/highlight.min.js"></script><script>hljs.highlightAll();</script><!-- Global site tag (gtag.js) - Google Analytics v4 --><script async src="https://www.googletagmanager.com/gtag/js?id=G-MCD87E7SF0"></script><script>window.dataLayer = window.dataLayer [];function gtag(){dataLayer.push(arguments);}gtag('js', new Date());gtag('config', 'G-MCD87E7SF0');</script><meta charset="utf-8" /><meta name="viewport" content="width=device-width">				

Assignment 2

GPT의 도움

웹 스크래핑이 뭐야?



웹 스크래핑(Web Scraping)은 자동화된 프로그램을 사용하여 웹사이트의 데이터를 추출하는 기술입니다. 웹 스크래핑을 통해 필요한 정보만을 정제하여 수집하고, 이를 데이터베이스나 파일로 저장할 수 있습니다.

웹 스크래핑의 기본 과정

1. 웹 페이지 접근: 스크립트나 프로그램이 특정 웹페이지에 접근합니다.
2. HTML 분석: HTML 구조를 분석하여 필요한 데이터가 위치한 태그, 클래스, 속성 등을 찾습니다.
3. 데이터 추출: 특정 태그나 속성에서 텍스트, 이미지 링크, 링크 등의 데이터를 추출합니다.
4. 데이터 저장: 추출된 데이터를 CSV, JSON, 데이터베이스 등에 저장해 활용합니다.

웹 스크래핑의 활용 예시

- 뉴스 사이트에서 최신 기사 제목과 링크를 수집하여 자동으로 리스트 생성
- 가격 비교 사이트에서 특정 제품의 가격 변동을 모니터링

<https://chatgpt.com/share/672074ee-5e44-8000-90ff-05c5e38d1497>