


| 카카오톡 봇 강좌 | >

[중급] (라이노 엔진) 자바스크립트 기초 강의 (5 - 1) - 더 나은 파싱 방법들 알아보기

 마른얼음 BOT 1:1 채팅

2019.07.27. 20:00 조회 532

💬 댓글 3

🔗 URL 복사

⋮

(5 - X 강좌들은 파싱을 더 잘 할 수 있도록 도와주는 강좌들이므로, 자신의 파싱 실력이 뛰어나다 하면 스킵하셔도 됩니다)

이번 시간에는 더 효율적이고 더 나은 HTML 파싱 방법들을 알아보겠습니다.
일단 HTML 파싱을 할 때 주로 태그를 먼저 꼼살(?)시켜두고 파싱하시는 분들이 많은데, 이렇게 하면 파싱 난이도가 매우 올라갈 수 있습니다.

이유를 설명해드리자면, 만약 자신이 날씨를 가져오고 싶는데 태그를 지웠을 때 날씨 정보들 사이에 고정된 값이 없으면 어디에서 끊고 어디를 바꿔야 할 지 모르죠. 하지만 태그는 고정되어 있는 경우가 많으며, 살짝 바뀌기도 하지만 태그명은 항상 고정되어 있기 때문에 자를 곳이 확실히 보이기 때문에 태그를 살려두고 파싱을 하는게 더 쉽습니다. 한 가지 더, 제발 띄어쓰기랑 줄바꿈 다 삭제하고 파싱하지마요. 가독성이 떨어지는건 말 할 필요도 없고 자신이 파싱할 때에도 가독성이 떨어져서 더 헛갈립니다.

그리고 여기서 더 나아가서 Jsoup을 활용하면, Utils.getWebText를 쓰는것보다 더 간단합니다. Utils.getWebText에서 split 서너번 하고 복잡한 정규식으로 replace 많이 하는것 보다 Jsoup에서 select 한두번과 text만으로 끝이기 때문입니다.

그리고 한 가지 말씀드릴게 있다면, Utils.getWebText 파싱방법에서 가끔 전체를 짜집기하면서 파싱하시는 분들이 많은데, 이러면 소스가 너무 복잡해집니다. 부분적으로 한개씩 자르고 변수에 담아서 최종정리하는게 훨 나아요!

이제부터 파싱하는 소스의 구조와 구체적으로 값을 뽑아오는 방법을 알려드리겠습니다.

- Utils.getWebText 방법
소스 구조)

```
var parse1 = Utils.getWebText(링크); //필요하면 split으로 자르기
//parse2, parse3으로 쪽쪽 이어나가기
/* 예를 들면
var parse2 = parse1.split(거시기)[0].split(뭐시기)[1];
var parse3 = parse1.split(알랄라)[0].split(랄랄라)[1];
이런식이죠 */
replier.reply(parse2+"\\n"+parse3...)
```

간단한 구조는 이렇습니다.
parse1 변수를 만들어서 거기에 HTML을 저장해놓고, 그 뒤 다른 변수(여기서는 parse2, parse3 등등)에 parse1을 구체적으로 잘라서 저장한 뒤, 변수들 모두 합쳐주면 됩니다. 반복적인 작업을 많이 하는 경우에는 (예 : 실검, 블로그글 목록, 카페글 목록, 주간날씨 등) 빈 배열을 만들고 for문에서 배열에 push하는 방식으로 하면 됩니다.

예외처리를 할 때에는 try - catch문으로 감싸서 예외처리 하거나, 어떤 특정 값이 null일때 예외처리를 하면 됩니다.

구체적으로 자를 때에는 어떻게 해야 할까요? 구체적으로 자를 때에는, 원하는 값 양 옆의 태그를 기준으로 잘라주시면 됩니다. 만약 양 옆의 태그가 앞에 또 있었어서 원하는 결과를 뽑기 어렵다면, 양 옆의 태그 앞부분을 잘라준 뒤 또 자르면 됩니다. 만약 사이사이에 태그가 있다면 그건 정규식으로 학살(?) 해버리자구요!

지금은 예제를 들면서 변수이름을 parse1, parse2 등으로 했지만, 사실 변수 이름은 값에 대한 걸로 해주는게 좋아요!
(예를 들면 게임이름 담아놓은 변수는 gameTitle, 랭킹 담아놓은 변수에는 ranking 등..)

- Jsoup 사용하는 방법
소스 구조)

```
var parse1 = org.jsoup.Jsoup.connect(링크) //필요하면 select
//다른 변수들 만들어서 쪽쪽쪽~
/* 예를 들면
var parse2 = parse1.select(불불불).text();
var parse3 = parse1.select(뽕뽕뽕).text();
이런식 */
replier.reply(parse2+"\n"+parse3...)
```

이런 구조로 하시면 됩니다. select는 만능이라(?) 더 말할 필요도 없겠네요. 마찬가지로 반복적인 작업시 for문 이용!
예외처리는 size를 체크해서 0일때 예외처리 하면 됩니다.

지금 봤을때 무슨 말인지 모르겠어도 앞으로 올라올 예제들을 보면서 이해하면 됩니다

강좌 끝
이번엔 설명을 살짝 못한 감이 있네요. 죄송합니다
다음번엔 예고했다시피 파싱 예제들을 올릴겁니다!
그럼 이만.



마른얼음 BOT님의 게시물 더보기 >

❤ 좋아요 4 💬 댓글 3

🔗 공유 | 신고

댓글 등록순 최신순 🔁

댓글알림 ☐



아미타

좋은글이네요. 배워갑니다.

2020.01.13. 22:08 답글쓰기



마른얼음 BOT 작성자

헐.. 좋은 글로 봐주셨다니 정말 감사드려요



2020.01.13. 22:10 답글쓰기



옛지오

얼음님 덕분에 많이 배워갑니다.~

2020.04.18. 17:32 답글쓰기

Hibot

댓글을 남겨보세요



등록

글쓰기

답글

목록

▲ TOP

'카카오톡 봇 강좌' 게시판 글

전체 [중급] 말머리 글

이 게시판 새글 구독하기 ☐

[초급] 봇 기초 강좌 1-1 [3]	렉스봇	2019.08.01.
[초급] json 강좌 [6]	렉스봇	2019.07.31.
[중급] (라이노 엔진) 자바스크립트 기초 강의 (5 - 1) - 더 나은 파싱 방법들 알아보기 [3]	마른얼음 BOT	2019.07.27.
[초급] 객체 생성하기 강좌 [7]	렉스봇	2019.07.27.
카카오톡봇 시작부터 지하철 실시간 정보까지 3편 자동응답 만들기(조건문) ☹️ [11]	사로로	2019.07.26.


1 2 3

전체보기

이 카페 인기글



메신저봇 삭제버튼 이렇게 안될까요?



특정방에서만 반응하는거..




기나무 곡괭이[노말] (오) 죄악의 문들 중 하나를

5000분의 1을 얻은 사람

eval 질문

Kiri

♡0 💬5



봇을쓰는데 메시지가 두번

간편 자동응답과 한글코딩을 메인으로 제공 하는 카톡봇앱 개발 예정

성빈

♡1 💬14

카페에 RPG가 많이 보이는 것 같군요

AlphaDo

♡1 💬3

예러를 알림으로 남게 해보자!

Lunar

♡2 💬4

가르치기 리로드..

Milk2

♡0 💬9

1 2 3 4