

[| 카카오톡 봇 강좌 | >](#)

## [초급] 파싱 시 Utils 와 Jsoup 의 장단점



doami2005 챗봇 창시자 1:1 채팅

2019.08.08. 04:31 조회 591

댓글 9 URL 복사 ⋮

제가 파싱할 때 주로 html에서 태그와 띄어쓰기를

```
replace(/<([^\>]+)>/g, "").replace(/[\\n\\s]{2,}/g, "\\n")
```

이 방법으로 싹 소멸시켜버리는데

결과 자체는 되게 깔끔하여 보기 쉬우며

split 을 쓰기 쉽다는 이점이 있습니다.

하지만, 특정 사이트들은 태그에 필요한 정보가 있어서 태그를 제거해버리면 정보가 사라집니다.

(태그란? 웹사이트 html에

<와 > 사이에 써있는 것

예) <div id=~~>

<div class~~>

등등)

그래서 파싱할 때 태그 제거는

미리 제거 한 뒤 결괏값을 보고,

필요한 정보가 쓰여있지 않다면 태그를 보면서 하시는 것을 추천드립니다.

태그 제거 전 —

SKT

📶 13% 🔋 오전 4:29

[← 전체보기](#)

```
...iduen= false <
```

```
<h3 class="blind">급상승 검색어 검색어</h3>
```

```
<div class="ah_roll_area
```

```
PM_CL_realtimeKeyword_rolling">
```

```
<ul class="ah_l">
```

```
<li class="ah_item"> <a href="#" class="ah_a" data-clk="lve.keyword"> <span class="ah_r">1</span> <span class="ah_k">위메프88특가</span>
```

</a> </li>

<li class="ah\_item"> <a href="#" class="ah\_a" data-clk="lve.keyword"> <span class="ah\_r">2</span> <span class="ah\_k">백종원 골목식당</span> </a> </li>

<li class="ah\_item"> <a href="#" class="ah\_a" data-clk="lve.keyword"> <span class="ah\_r">3</span> <span class="ah\_k">김민주</span> </a> </li>

<li class="ah\_item"> <a href="#" class="ah\_a" data-clk="lve.keyword"> <span class="ah\_r">4</span> <span class="ah\_k">입추</span> </a> </li>

<li class="ah\_item"> <a href="#" class="ah\_a" data-clk="lve.keyword"> <span class="ah\_r">5</span> <span class="ah\_k">골목식당</span> </a> </li>

<li class="ah\_item"> <a href="#" class="ah\_a" data-clk="lve.keyword"> <span class="ah\_r">6</span> <span class="ah\_k">로켓펀치</span> </a> </li>

<li class="ah\_item"> <a href="#" class="ah\_a" data-clk="lve.keyword"> <span class="ah\_r">7</span> <span class="ah\_k">송병준</span> </a> </li>

<li class="ah\_item"> <a href="#" class="ah\_a" data-clk="lve.keyword"> <span class="ah\_r">8</span> <span class="ah\_k">김민준</span> </a> </li>

태그 제거 후 —

SKT 

   13%  오전 4:29

← 전체보기



π—  
증권(금융)  
부동산  
지도  
영화  
음악  
한

색  
만화 / 웹툰  
더보기  
급상승 검색어 검색어  
1 위메프88특가  
2 백종원 골목식당  
3 김민주  
4 입추  
5 골목식당  
6 송병준  
7 로켓펀치  
8 기생충 해석  
9 케빈 나  
10 기문둔갑  
11 한끼줍쇼  
12 이진혁  
13 워너원  
14 김희철 모모  
15 포방터 돈까스  
16 이대 백반집  
17 쫄지  
18 봉오동 전투  
19 아드레날린24  
20 주옥순  
급상승 검색어  
DataLab.급상승 트래킹  
1~10의

네이버는 필요한 정보가 태그에 저장되어 있지 않아  
태그 제거 시에 보기가 매우 깔끔해집니다

대표적인 정규식 말고도 필요에 따라 본인이 개발해서 쓰는것이 중요합니다

/\* 도미의 강좌 시리즈 \*/

[Tip]

\*eval 의 위험성과 사용법\*

- <https://cafe.naver.com/nameyee/4341>

\*파싱 시 Utils 와 Jsoup 의 장단점\*

- <https://cafe.naver.com/nameyee/4342>



doami2005님의 게시글 더보기 >

❤ 좋아요 2 💬 댓글 9

🔗 공유 | 신고

댓글 등록순 최신순 ↻

댓글알림 ☐



doami2005 작성자

파싱 위주 소스 요청 받습니다..

2019.08.08. 04:35 답글 쓰기



성빈

이번주까지 다 나오는 날씨요  
네이버 파싱하시면 됨

2019.08.08. 08:27 답글쓰기



사로로

서해선 실시간이요

2019.08.08. 10:56 답글쓰기



doami2005 작성자

이사람들...이상해

2019.08.08. 17:21 답글쓰기



jihun0302

아 서울지역 이외지하철이요

2019.08.08. 17:21 답글쓰기



릭스봇

성빈님의 남은 수능 일수 알려주기

2019.08.08. 21:17 답글쓰기



목삿

<http://rawris-am.ekr.or.kr/wrms/>  
저수율 검색이요!!

2019.08.09. 16:48 답글쓰기



doami2005 작성자

사이트에 문제가 있다네요

2019.08.09. 16:59 답글쓰기



목삿

[https://www.water.or.kr/realtime/sub01/sub01/dam/hydr.do?s\\_mid=1323&seq=1408&p\\_group\\_seq=1407&menu\\_mode=3&damcd=4007110#tab-contentArea](https://www.water.or.kr/realtime/sub01/sub01/dam/hydr.do?s_mid=1323&seq=1408&p_group_seq=1407&menu_mode=3&damcd=4007110#tab-contentArea)

2019.08.09. 18:46 답글쓰기

Hibot

댓글을 남겨보세요



등록

글쓰기

답글

목록

▲ TOP

## '| 카카오톡 봇 강좌 |' 게시판 글

전체 [초급] 말머리 글

이 게시판 새글 구독하기 ☐

윽챗똥으시려는 분들을 위한 자그마한 팁 [28]	아이곰	2019.08.15.
[초급] 오류를 잡아주는 try문 사용법 [7]	doami2005	2019.08.15.
[초급] 파싱 시 Utils 와 Jsoup 의 장단점 ☹️ [9]	doami2005	2019.08.08.
[초급] eval의 위험성/사용법 ☹️ [11]	doami2005	2019.08.08.
[초급] 봇 기초 강좌 1-1 [3]	럭스봇	2019.08.01.


1 2 3

전체보기

## 이 카페 인기글

반가워요.

천방지축하연  
♡0 💬4



전에 말했던 마피아 계열 게임 소스 올렸습  
니다.  
카카오톡 단순 자동응답봇 파일  
ZUMP  
도움료  
♡0 💬2



시간별/날짜별 날씨 구현

Dark Tornado  
요즘 나타나는 새로운 홍보 수법 중 하나가  
Dark Tornado  
♡0 💬7

중수가 초보에게 하는 강좌 [ 2 ]

하하하하  
♡1 💬11

가르치기 리로드..

Milk2  
♡0 💬9

[ 치\*\*\*, 한\*\* ] 님 강제 탈퇴

AlphaDo  
♡1 💬2

JPG 개발 일지 #7

재승  
♡0 💬4

1 2 3 4