

| 카카오톡 봇 강좌 | >

## [중급] 크롤링/파싱 예제 (한강 수온 예시)



**Dark Tornado** 카페스텝 1:1 채팅

2020.02.16. 10:29 조회 353

댓글 8 URL 복사

(>\_<)

출처 블로그>Dark Tornado | Dark Tornado

원문 <http://blog.naver.com/dt3141592/221810865748>

크롤링/파싱 예제를 보고 따라만 하거나 복붙하는 사람들이 있습니다.

이건 수학시간에  $1 + 1 = 2$  라는 내용을 보고,

$1 + 1$ 은 2구나 라고 따라하거나 외우고 끝난겁니다.

+가 뭔지 모르는 상태로 넘어가는거예요.

그러니, 자신이 만든 결과물에  $1 + 1$ 은 있는데,  $1 + 2$ 는 없고, 만들지도 못하는 것이지요.

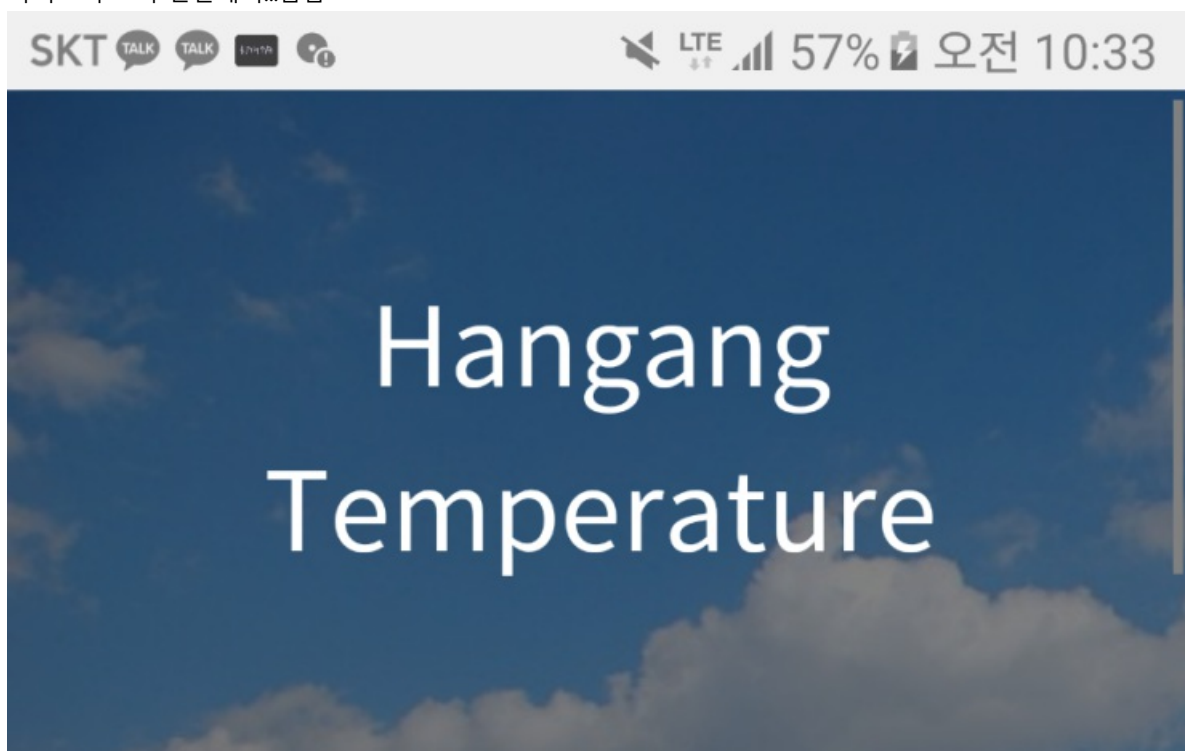
ex) 날씨 파싱은 봇에 넣었으면서, 비트코인 파싱은 못하는 경우. 사실, 파싱이 무엇인지 모르거나, 본인 봇에 있는 소스가 파싱을 하는지도 모르는 경우도 있...읍읍

하여튼, 대강 인터넷에서 무언가를 긁어오는 것을 크롤링, 긁어온 결과를 지지고 볶는걸 파싱이라고 한다고 알고 계시면 됩니다.

방법은 간단합니다.

1. 내장 메소드 또는 Jsoup 라이브러리를 이용하여 원하는 웹사이트의 내용을 가지고온다.
2. 원하는 부분만 남기거나 뽑는다.
3. 그 부분에 이것저것 붙이거나, 그냥 끝.

이 게시글에서는 한강 수온을 가지고 오는 기능을 만들 예정이고, 한강 수온은 이 사이트에서 확인할 수 있습니다. 사실, 사이트 구조가 간단해서...읍읍



현재 한강물의 온도는

💧 4.3도

[2020-02-02 00:30:03] 에 측정된 자료.  
WallPaper Engine에서도 한강물 온도계를 만  
나보세요!

©WPWS 2018

#### 1. 내장 메소드 사용해서 직접 지지고 볶기

Utils.getWebText(url);

해당 url의 HTML 소스를 읽어옵니다.

HTML이 뭐냐구요? 웹사이트 만들 때 사용하는 마크업 언어입니다.

하여튼, 저 메소드는 젤브봇(보라봇), 채팅 자동응답 봇(초록봇), 메신저봇, 메신저봇, Nusty 내장 챗봇 등에 있습니다.



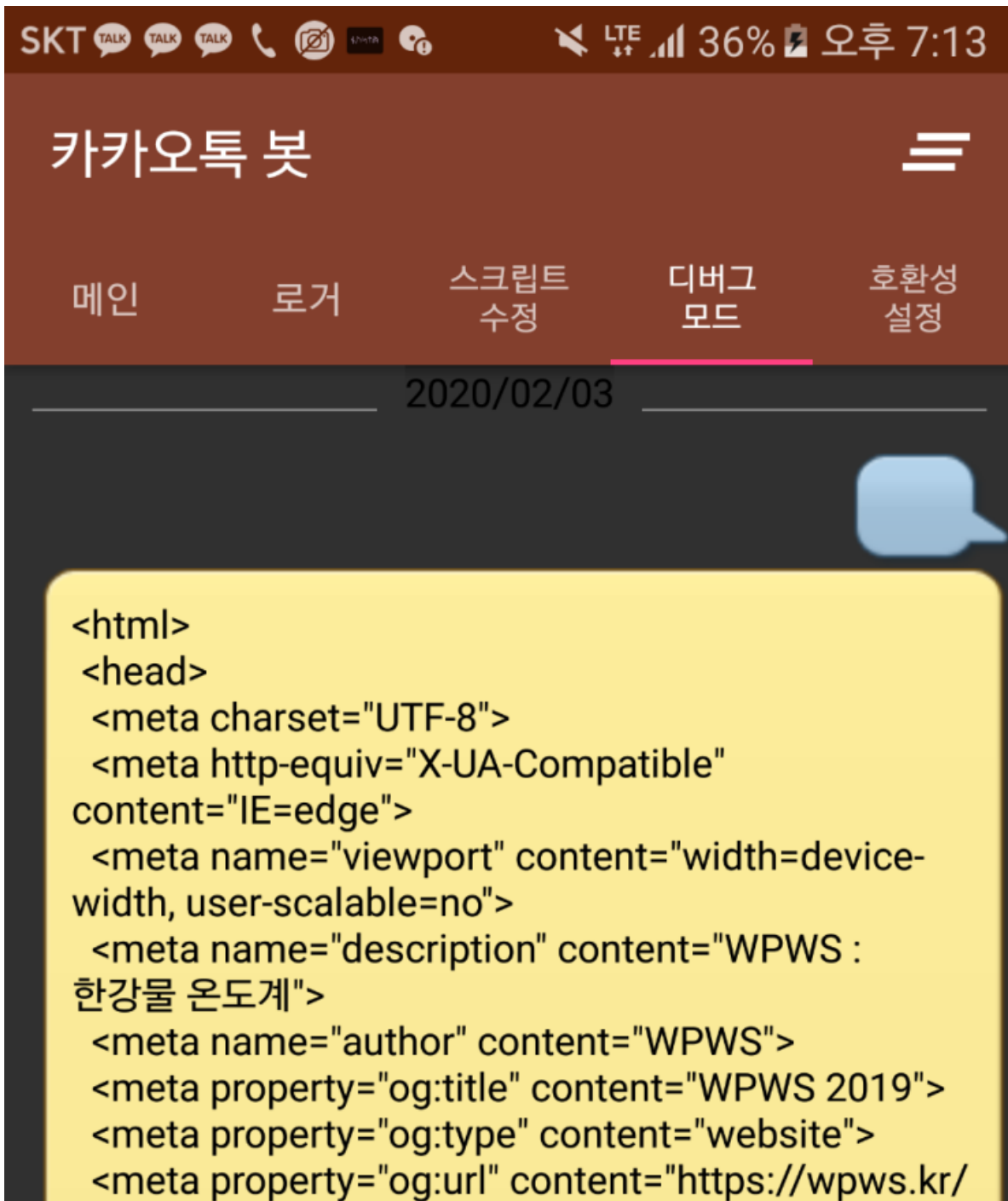
WebMasterLite



```
1 function response(room, msg, sender,  
  isGroupChat, replier) {  
2   var data = Utils.getWebText("https://
```

```
3 www.wpws.kr/hangang/");
4 replier.reply(data);
5
6
7 }
8
9
10
11
```

일단, 굵어온 내용을 출력해봅시다.



```
hangang">
```

```
<meta property="og:description" content="WPWS :  
한강물 온도계">
```

```
<meta name="twitter:title" content="WPWS :  
WingPlanet Web Service">
```

```
<meta name="twitter:site" content="https://  
wpws.kr/hangang">
```

메세지를 입력하세요



프리메라 슈퍼블랙씨드 콜드드랍

얼리-안티에이징을 위한 프리메라의 야심작 Apmall

자세히 알아보기

원하는 부분인 한강 수온이 여기에 있군요.

SKT TALK TALK



LTE



58% 오전 10:39

Jota Text Editor\*



더보기

```
<div id= cover >
```

```
<div id="content">
```

```
<p id="title">Hangang Temperature</p>
```

```
<br>
```

```
<br>
```

```
<p id="subtitle">현재 한강물의 온도는</  
p>
```

```
<p id="temp"><i class="xi-tint"></  
i>4.3도</p>
```

```
<br>
```

```
<br>
```

```
<p id="foo2"> [2020-02-02 00:30:03] 에  
측정된 자료. <br><a href="../sat"  
style="color:white">여기</a>에서 등급컷도  
알아보는거 어때요?<br> <br>©WPWS 2018
```

```

</p>
</div>
</div>
<script src="./g.js"
type="c74cc20157ab82b8ff741194-text/
javascript"></script>
<script src="https://ajax.cloudflare.com/
cdn-cgi/scripts/7089c43e/cloudflare-
static/rocket-loader.min.js" data-cf-

```

Save Undo Redo Copy Cut Paste History Quit

일단 정규식을 통해서 모든 태그들을 삭제해봅시다.

<이렇게> 생긴 것들이 태그입니다.

초록색은 Utils.removeTags(); 메소드 사용해도 상관\*



WebMasterLite



```

1 function response(room, msg, sender,
  isGroupChat, replier) {
2   var data = Utils.getWebText("https://
  www.wpws.kr/hangang/");
3   data = data.replace(/(<([^\>]+)>)/g, "");
4
5   replier.reply(data);
6
7
8 }

```

보

SKT TALK TALK



LTE



58%



오전

10:40

카카오톡 봇





메인

로거

스크립트  
수정

디버그  
모드

호환성  
설정

WPWS : 한강 수온 체크기

Hangang Temperature

현재 한강물의 온도는  
4.3도

[2020-02-02 00:30:03] 에 측정된 자료. WallPaper  
Engine에서도 한강물 온도계를 만나보세요! ©WPWS  
2018

메세지를 입력하세요



저 부분만 남기고 다 날릴겁니다.

SKT TALK TALK 4Dnet



LTE



58%



오전 10:40



## Hangang Temperature

현재 한강물의 온도는

4.3도

---



[2020-02-02 00:30:03] 에 측정된 자료.  
WallPaper Engine에서도 한강물 온도계를  
만나보세요! ©WPWS 2018

Save

Undo

Redo

Copy

Cut

Paste

History

Quit

.split(); 메소드는 해당 문자열을 특정 문자열 단위로 잘라줍니다(?).

```
var a = "나는 집에 가고싶다".split(" ");
```

를 한다면, 띄어쓰기 단위로 문자열이 잘려서,

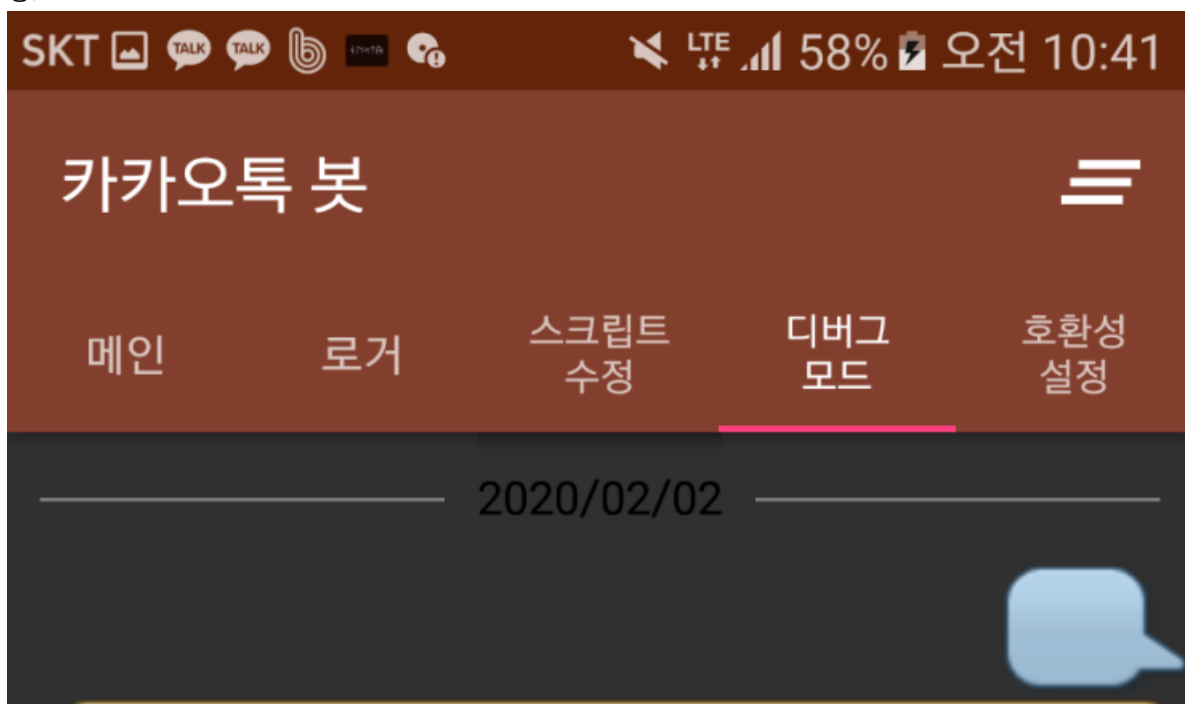
```
["나는", "집에", "가고싶다"]
```

가 됩니다. 배열이네요.

하어튼, 자른 결과 중 1번째 녀석만 남길겁니다. 0부터 시작해요

```
1 function response(room, msg, sender,
2   isGroupChat, replier) {
3   var data = Utils.getWebText("https://
4     www.wpws.kr/hangang/");
5   data = data.replace(/(<([>]+)>)/g, "");
6   data = data.split("온도는");
7   data = data[1];
8
9   replier.reply(data);
10
11 }
```

땡.





4.3도

[2020-02-02 00:30:03] 에 측정된 자료. WallPaper Engine에서도 한강물 온도계를 만나보세요! ©WPWS 2018

메세지를 입력하세요



도를 기준으로 다시 한 번 자르고, 0번째 녀석만 남기면,



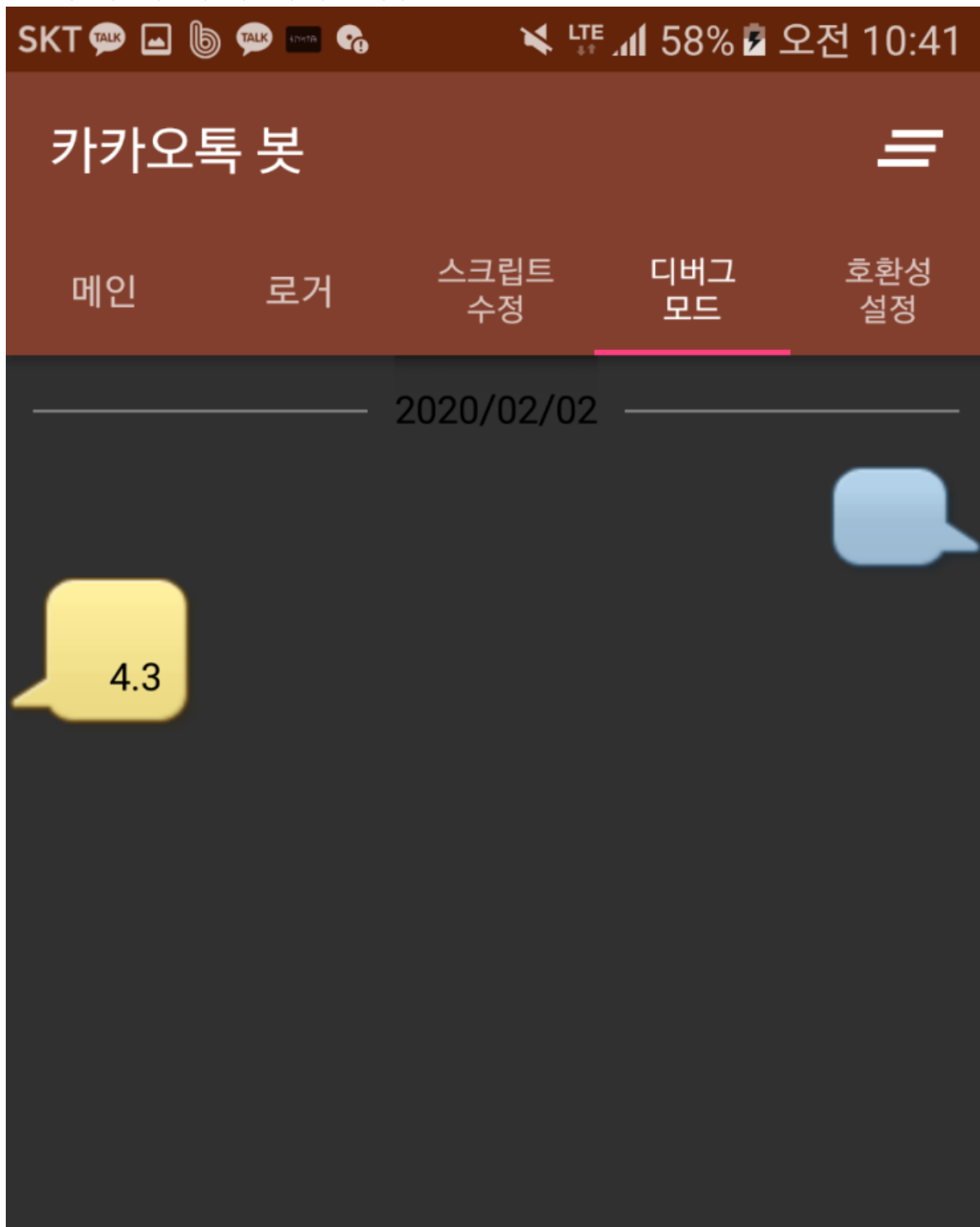
WebMasterLite



```
1 function response(room, msg, sender,  
  isGroupChat, replier) {  
2   var data = Utils.getText("https://  
   www.wpws.kr/hangang/");  
3   data = data.replace(/(<([^\>]+)>)/g, "");  
4   data = data.split("온도는");  
5   data = data[1];
```

```
6 data = data.split("도");
7 data = data[0];
8
9 replier.reply(data);
10
11
12 }
13
```

뽕. 원하는 부분이 공백들과 함께 나왔습니다.



메세지를 입력하세요



공백들은 .trim(); 메소드로 조져주면 됩니다.

```
1 function response(room, msg, sender,  
2 isGroupChat, replier) {  
3   var data = Utils.getWebText("https://  
4   www.wpws.kr/hangang/");  
5   data = data.replace(/(<([>]+)>)/g, "");  
6   data = data.split("온도는");  
7   data = data[1];  
8   data = data.split("도");  
9   data = data[0];  
10  data = data.trim();  
11  
12  replier.reply(data);  
13 }  
14
```

땡. 온도 부분만 추출 완료.

SKT   100%

 LTE  58%  오전 10:42

카카오톡 봇



메인

로거

스크립트  
수정

디버그  
모드

호환성  
설정

2020/02/02

4.3

메세지를 입력하세요



이제, 이것저것 붙이면 됩니다.



WebMasterLite

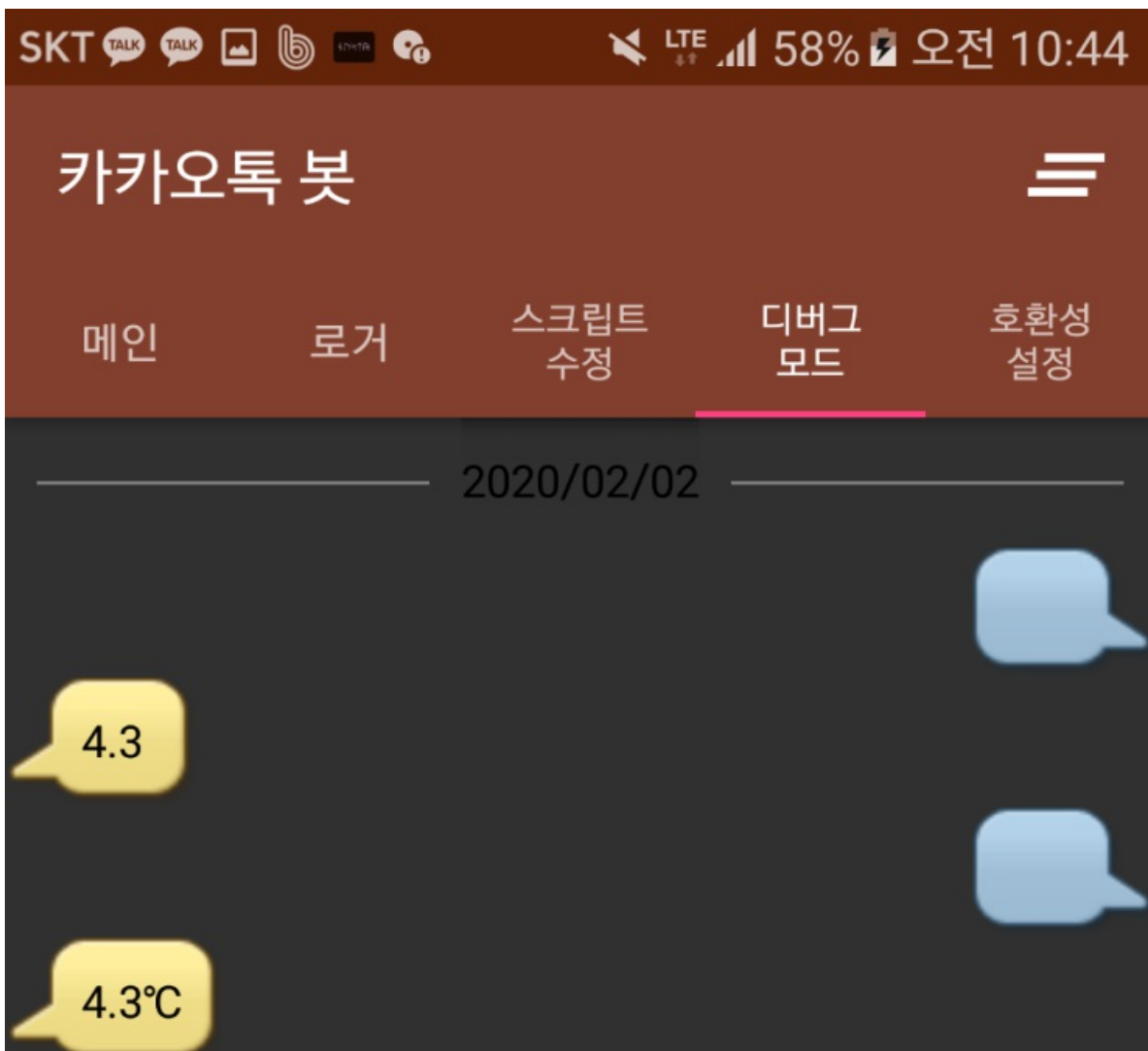


```

1 function response(room, msg, sender,
  isGroupChat, replier) {
2   var data = Utils.getWebText("https://
  www.wpws.kr/hangang/");
3   data = data.replace(/(<([>]+)>)/g, "");
4   data = data.split("온도는")[1];
5   data = data.split("도")[0];
6   data = data.trim();
7
8   replier.reply(data + "°C");
9
10  }
11

```

앙



메세지를 입력하세요



if도 슬쩍 넣어주면

```
1 function response(room, msg, sender,  
  isGroupChat, replier) {  
2   if (msg == "/한강") {  
3     var data = Utils.getWebText("https://  
  www.wpws.kr/hangang/");  
4     data = data.replace(/(<([^\>]+)>)/g, "");  
5     data = data.split("온도는")[1];  
6     data = data.split("도")[0];  
7     data = data.trim();  
8     replier.reply(data + "°C");  
9   }  
10 }  
11
```

완성.



# 카카오톡 봇



메인

로거

스크립트  
수정

디버그  
모드

호환성  
설정

2020/02/02

4.3

4.3°C

4.3°C

4.3°C

/한강

앙 너스띠

/한강

메세지를 입력하세요



전체 소스 :

```
response.js
1 function response(room, msg, sender, isGroupChat, replier) {
2   if (msg == "/한강") {
3     var data = Utils.getText("https://www.wpws.kr/hangang/");
4     data = data.replace(/(<([^\>]+)>)/g, "");
5     data = data.split("온도는")[1];
6     data = data.split("도")[0];
7     data = data.trim();
8     replier.reply(data + "°C");
9   }
10 }
11
```

## 2. Jsoup 라이브러리 사용

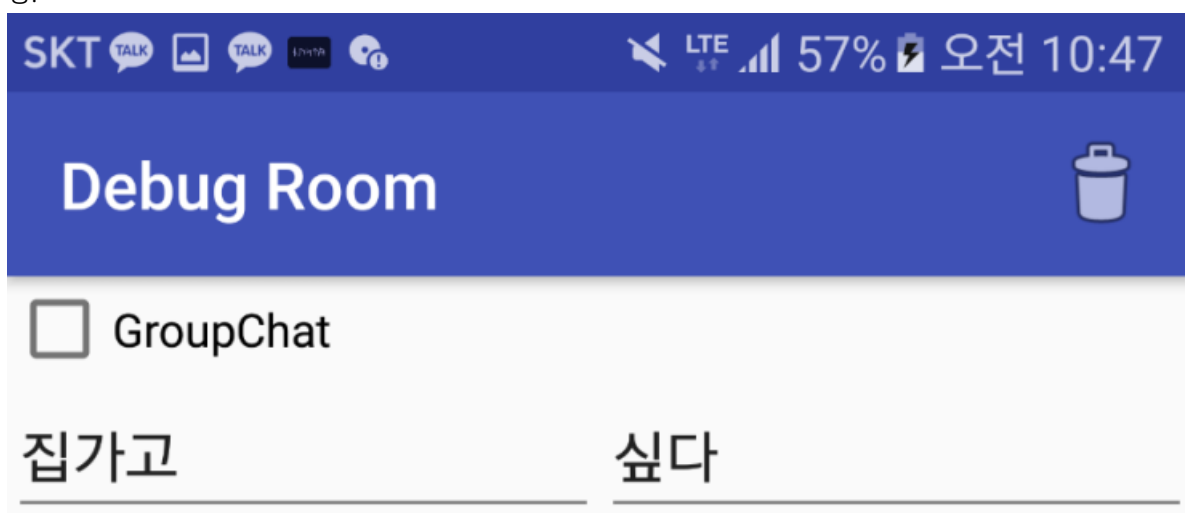
웹사이트 긁는 용도로 만들어진 Jsoup이라는 라이브러리가 있습니다.

채팅 자동응답 봇(초록봇), 메신저봇, 메신저봇, Nusty 내장 챗봇에 내장되어 있으며, 젤브봇(보라봇)은 기억이 가물가물...

하여튼, 사이트 내용을 전부 긁어오면,

```
1 function response(room, msg, sender,
2   isGroupChat, replier, ImageDB) {
3   var data = org.jsoup.Jsoup.connect("https://
4     www.wpws.kr/hangang/").get();
5   replier.reply(data);
6 }
7
```

뽕.



싶다

냥

```
<html>
<head>
  <meta charset="UTF-8">
  <meta http-equiv="X-UA-Compatible"
content="IE=edge">
  <meta name="viewport" content="width=device-
width, user-scalable=no">
  <meta name="description" content="WPWS :
한강물 온도계">
  <meta name="author" content="WPWS">
  <meta property="og:title" content="WPWS
2019">
  <meta property="og:type" content="website">
  <meta property="og:url" content="https://
wpws.kr/hangang">
  <meta property="og:description"
content="WPWS : 한강물 온도계">
  <meta name="twitter:title" content="WPWS :
WingPlanet Web Service">
  <meta name="twitter:site" content="https://
wpws.kr/hangang">
  <meta name="twitter:description"
```

Enter message

SEND

원하는 부분인 온도는 이 곳에 있습니다.

SKT TALK 57% 오전 10:47

Jota Text Editor\*



더보기

```
<p id="title">Hangang Temperature</p>
<br>
```

```

<br>
<p id="subtitle">현재 한강물의 온도는</
p>
<p id="temp"><i class="xi-tint"></
i>4.3도</p>
<br>
<br>
<p id="foo2"> [2020-02-02 00:30:03] 에
측정된 자료. <br><a href="../sat"
style="color:white">여기</a>에서 등급컷도
알아보는거 어때요?<br> <br>©WPWS 2018
</p>
</div>
</div>
<script src="./g.js"
type="14f60963b96c92baf51e37c5-text/
javascript"></script>
<script src="https://ajax.cloudflare.com/
cdn-cgi/scripts/7089c43e/cloudflare-
static/rocket-loader.min.js" data-cf-
settings="14f60963b96c92baf51e37c5-
|49" defer></script>
" .

```

Save Undo Redo Copy Cut Paste History Quit

id가 temp인 <p> 태그 내부에 온도가 포함되어 있군요.

```

<br>
<br>
<p id="subtitle">현재 한강물의 온도는</
p>
<p id="temp"><i class="xi-tint"></

```

i>4.3도</p>

<br>

<br>


<p id="foo2"> [2020-02-02 00:30:03] 에  
측정된 자료. <br><a href="../sat"  
style="color:white">여기</a>에서 등급컷도  
알아보는거 어때요?<br> <br>©WPWS 2018  
</p>

그럼, .select(); 메소드를 통해 저 부분만 뽑아줍니다.

```
1 function response(room, msg, sender,  
  isGroupChat, replier, ImageDB) {  
2   var data = org.jsoup.Jsoup.connect("https://  
  www.wpws.kr/hangang/").get();  
3   data = data.select("p[id=temp]");  
4   replier.reply(data);  
5 }  
6  
7
```

뽕.

SKT TALK TALK b 11:11 58% 오전 10:51

Debug Room 

☐ GroupChat

집가고 

싫다

싫다

냥

<p id="temp"><i class="xi-tint"></i>4.3도</p>

Enter message

SEND

긏어온 결과에 .text(); 메소드를 사용하면,

```
1 function response(room, msg, sender,  
  isGroupChat, replier, ImageDB) {  
2   var data = org.jsoup.Jsoup.connect("https://  
   www.wpws.kr/hangang/").get();  
3   data = data.select("p[id=temp]").text();  
4   replier.reply(data);  
5 }
```

태그들의 사라지고 내용물(?)만 나옵니다



## Debug Room



☐ GroupChat

집가고

싶다

싶다

냐앙~☆

4.3도

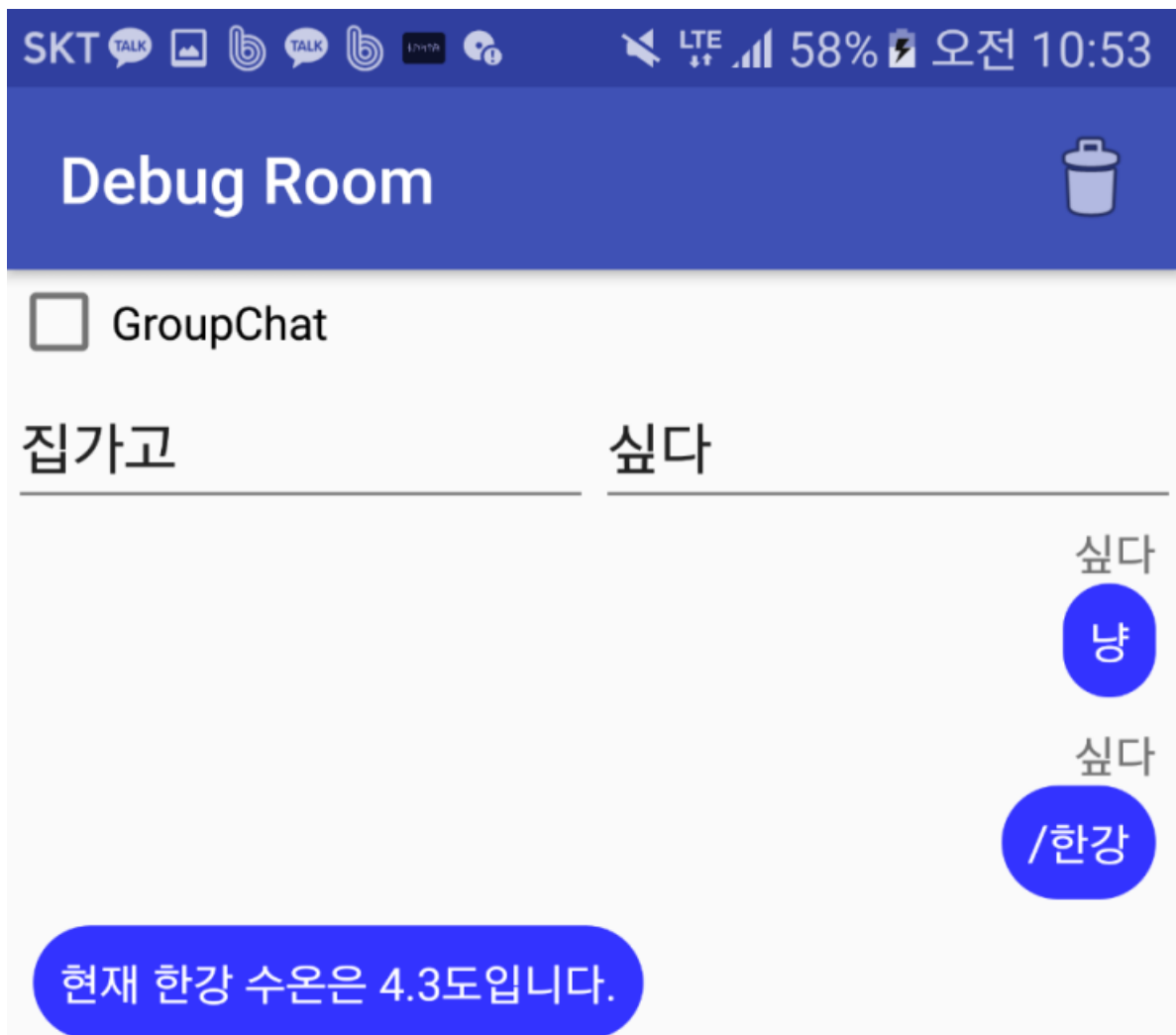
Enter message

SEND

if 같은거 슬쩍 넣어주면.

```
WebMasterLite
1 function response(room, msg, sender,
  isGroupChat, replier, ImageDB) {
2   if (msg == "/한강") {
3     var data =
org.jsoup.Jsoup.connect("https://
www.wpws.kr/hangang/").get();
4     data = data.select("p[id=temp]").text();
5     replier.reply("현재 한강 수온은 " + data +
"입니다.");
6   }
7 }
8
```

끝



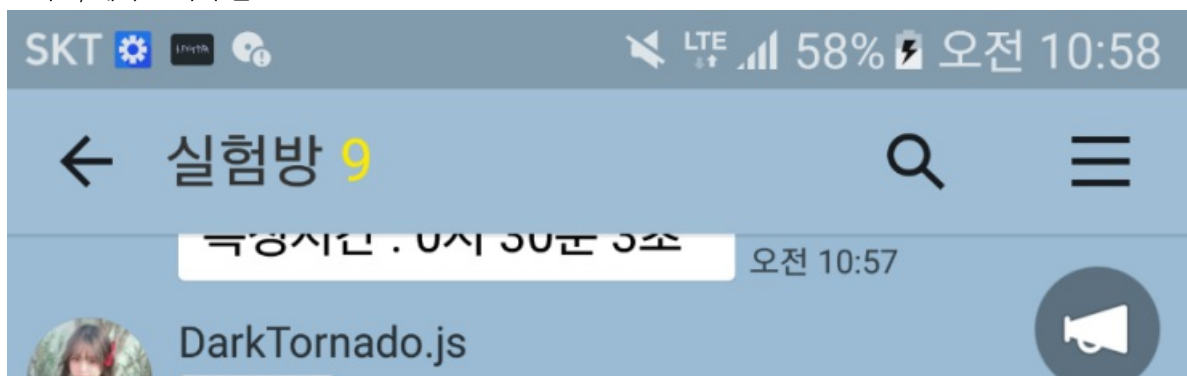
Enter message

SEND

전체 소스

```
<I> | 액티비티.js
1 function response(room, msg, sender, isGroupChat, replier, ImageDB) {
2   if (msg == "/한강") {
3     var data = org.jsoup.Jsoup.connect("https://www.wpws.kr/hangang/").get();
4     data = data.select("p[id=temp]").text();
5     replier.reply("현재 한강 수온은 " + data + "입니다.");
6   }
7 }
8
9 |
```

크리고, 에어로 이식 완료





/한강

5

오전 10:57



[AI] 닥땅.js

한강 수온 : 4.3°C  
측정일 : 2020년 2월 2일  
측정시간 : 0시 30분 3초

5

오전 10:57



[Bot] EA (에아)

현재 한강 수온은 4.3도입니다.

6

오전 10:57



DarkTornado.js

ㅈㅈ하하하하



7

오전 10:57

8

오전 10:58

/한강



[Bot] EA (에아)

현재 한강 수온은 4.3도입니다.

7

오전 10:58



Dark Tornado님의 게시글 더보기 >



좋아요 4



댓글 8



공유 | 신고

댓글

등록순

최신순



댓글알림



hyun123





보기에는 쉬운것 같지만 직접 해보면.. 어려울 것 같ㄸr..

2020.02.16. 11:26 답글 쓰기



같은

많은 도움이 되었습니다. 감사합니다.

2020.02.16. 11:40 답글 쓰기

삭제된 댓글입니다.



**Dark Tornado** 작성자

영?

2020.02.16. 12:21 답글 쓰기



사로로

너무 어려워요

2020.02.16. 15:19 답글 쓰기



인디벨

어디서 많이 본 듯한...음

2020.02.16. 17:05 답글 쓰기



**Leesy006**

파싱 강좌라니..감사합니다!!



2020.02.24. 22:28 답글 쓰기



로봇인데이

와.. 잘 읽었습니다. 넘 좋네요

2020.02.25. 12:18 답글 쓰기



검정쿠페

감사합니다

2020.05.18. 23:58 답글 쓰기

**Hibot**

댓글을 남겨보세요



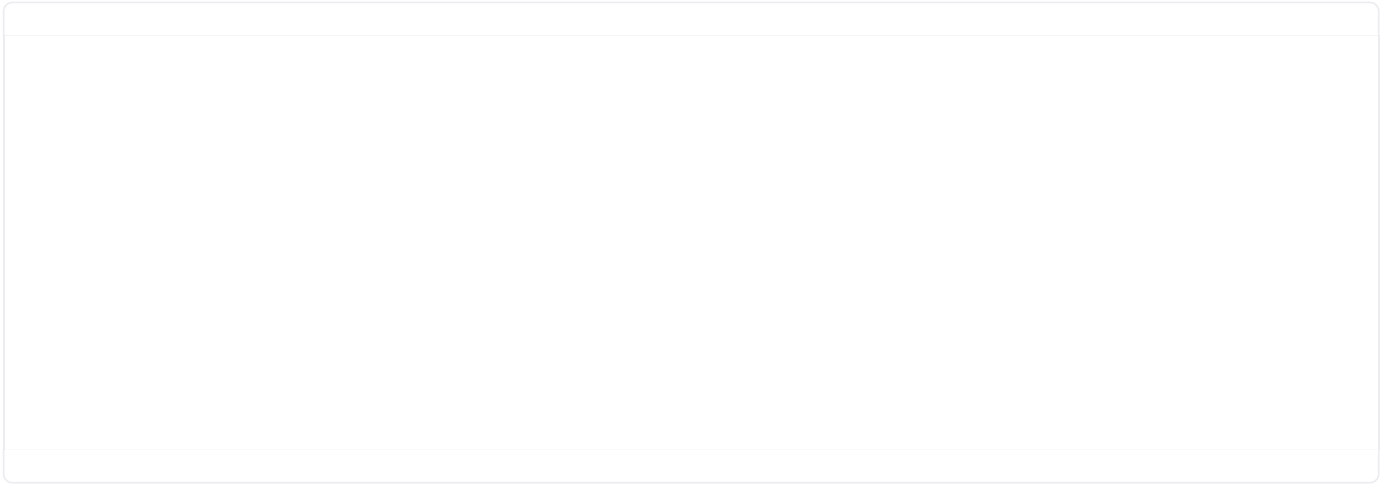
등록

글쓰기

답글

목록

▲ TOP



'| 카카오톡 봇 강좌 |' 게시판 글

전체 [중급] 말머리 글 이 게시판 새글 구독하기 ☐

RPG 제작기 0편 객체지향이 뭐지?(개판) [3]	사로로	2020.02.21.
[기타] [초급] Chat Bot Simulator 새 스크립트 시 나오는 기본 값 수정 [1]	대기만성	2020.02.18.
[중급] 크롤링/파싱 예제 (한강 수온 예시) [8]	Dark Tornado	2020.02.16.
Array.prototype.sort [4]	화베	2020.02.13.
Node.js Setting - Install Library - Express [19]	CodeMonkey	2020.02.11.

1 2 3 전체보기

이 카페 인기글

예전에 어떤분이 카톡봇을 이용하여 카페에  
게시글을 올렸었는데

리페  
♡0 💬1

함수를 입력받아 그래프 그리기

Steve28  
♡2 💬4

자동학습 봇 특정 말만 무시하기?

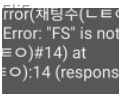
헤히  
♡0 💬4

안녕하세요!

트위터봇  
♡0 💬6



링크를 포함한 글 홍보시 링크 분석(?)하기



채팅순위봇이 갑자기 안되네요..ㅠ

공식 카카오톡 윙챗

Lunar  
♡0 💬9



혹시 여기서 변수를 저장하는 부분이 어딘지  
아시는분 계신가요??  
오픈이발 괴롭히기(메봇R)

Milk2  
ERROR  
♡0 💬7

1 2 3 4 5