

| 카카오톡 봇 강좌 | >

# [중급] 파싱 주소 찾기 (파싱이 안될 때; ajax; 비동기)

아미타

챗봇 중수

1:1 채팅

2020.01.26. 19:38

조회 446

...
댓글 13
URL 복사

## 강좌할 주제

=> 파싱이 안 되는 페이지 파싱하기 - ajax 페이지(비동기) 파싱하기

## 강좌 내용

=>

\*항상 이 카페에서 도움만 받아서..  
배운 내용을 종합해서 저도 누군가에게 도움이 되었으면 해서 글을 작성합니다.

우리가 알고있는 Utils.getWebText 혹은 org.jsoup.Jsoup 등을 이용해도 파싱할 수 없는 페이지도 있습니다.  
제가 이것 때문에 좀 고생을 해서.. 고생하시는 분들을 위해 도움이 되고자 글을 쓰게 됐습니다.  
아마 기존 강좌글에 중복된 개념도 있을 수 있는데요, 정리+끌어올리기 개념으로 너그려이 봐주셨으면 합니다.

왜 파싱이 안되는 페이지가 존재할까요..  
조금 간단히 말씀드리면, 그런 페이지들은 보여지는 부분의 html 태그로 구현되어 있지 않고 script를 이용해서 해당 내용을 구성하게 됩니다.  
그런 페이지들을 어떻게 가져올 지에 대해 글을 써봅니다.

오늘은 네이버 증권 모바일페이지를 예제로 가져왔습니다.  
<https://m.stock.naver.com/item/main.nhn#/stocks/005930/total>

## 1. 파싱할 페이지 확인



날짜	외국인 보유율	기관	개인	증가	전일비
01.23.	-2,262,365 -57.16%	+1,906,336	+4,031,573	60,800	▼-1,500
01.22.	+123,499 57.19%	+1,649,008	-1,731,848	62,300	▲900

네이버 증권의 Mobile페이지입니다.

우리는 여기서 아래에 있는 투자자별 매매동향을 파싱하겠습니다.

```

<div class="goto_area">
  <a href="intent://addshortcut?
url=https://m.stock.naver.com&icon=https://ssl.static.net/static/nfinance/ic
o/2018_android_192x192_xoohedi.png&title=9F83843A448E3303648F862884+9FC3A639
08FA836380?serviceCode=stock&version=7#Intent:scheme=naversearchapp;action=an
droid.intent.action.VIEW;category=android.intent.category.BROWSABLE;package=c
om.nhn.android.search;end" onclick="nclk(this, 'fot.shortcut', '', '');">
class="goto"><span class="app_stock">네이버 증권</span><em>네이버앱의</em>
<strong>증권</strong> 홈화면에 바로가기 추가 <span class="ico_add"></span>
</a>
</div>
</footer>
<script type="text/javascript">
!function(t,o){function n(t){new e(t)}function e(t)
{this._init(t);e.prototype={constructor:e,_init:function(t)
{this._aUrlToBeLoad=t._aUrlToBeLoad,this._sCharSet=t._sCharSet||"utf-
8",this._sId=t._sId||"",this._fnCallback=t._fnCallback,this._fnOnLoadScript=thi
s._getFnOnLoadScript(),this._oModel=new
e.Model(this._aUrlToBeLoad,this._oView=new
e.View).setCharSet(this._sCharSet).setId(this._sId),this._oHandlerManager=e.H
andlerManagerFactory.getHandler(),this._load(),_load:function(){return
this._oModel.setScriptUrlToBeLoad(),!0===this._oModel.isAlreadyLoaded()?
(this._onLoadScriptForAlreadyLoaded(),o):
(this._oView.loadScript({sUrlToBeLoad:this._oModel.getCurrentLoadingUrl(),oHa
ndlerManager:this._oHandlerManager,fnOnLoadCallback:this._fnOnLoadScript}),th
is._oModel.addToLoadingInfo(this._oView.getLoadingScriptEl(),o)},_onLoadScri
ptForAlreadyLoaded:function()
{this._oModel.updateDataForAlreadyLoaded(),this._onAfterLoad(),_getFnOnLoadS
cript:function(){var t=this;return function()
{t._oModel.updateDataWithLoadedScript(this,t._onAfterLoad())},_onAfterLoad:f
unction(){return!1===this._oModel.isLoadedAllScript()?(this._load(),o):
(this._fireCallback(),o)},_fireCallback:function(){var
t=this._fnCallback;"function"==typeof t&&t()}}.e.View=function(t)
{this._init(t);e.View.prototype={constructor:e.View,_init:function()
{this._sCharSet="utf-
8",this._sId="",this._eLoadingScript=null},setCharSet:function(t){return
this._sCharSet=t,this._setId:function(t){return
this._sId=t,this._getLoadingScriptEl:function(){return
this._eLoadingScript},loadScript:function(t)
{this._makeScriptElToBeLoad(t),this._appendScriptForLoad(),_appendScriptForL
oad:function(){var t=document.getElementsByTagName("head")
[0];t.appendChild(this._eLoadingScript),_makeScriptElToBeLoad:function(t)
{var
o=t._oHandlerManager,n=document.createElement("script");n.setAttribute("type",
"text/javascript"),n.setAttribute("charset",this._sCharSet),n.setAttribute("i
d",this._sId),n.setAttribute("src",t._sUrlToBeLoad),o.attachOnLoadToScript(n,t

```

파싱을 하고싶은데..소스 보기를 하면 원하는 부분에 저런 스크립트만 있지 html 엘리먼트가 없는 것을 확인할 수 있습니다. 이런 상태에서는 getWebText나 Jsoup으로는 파싱하기 불가능합니다.

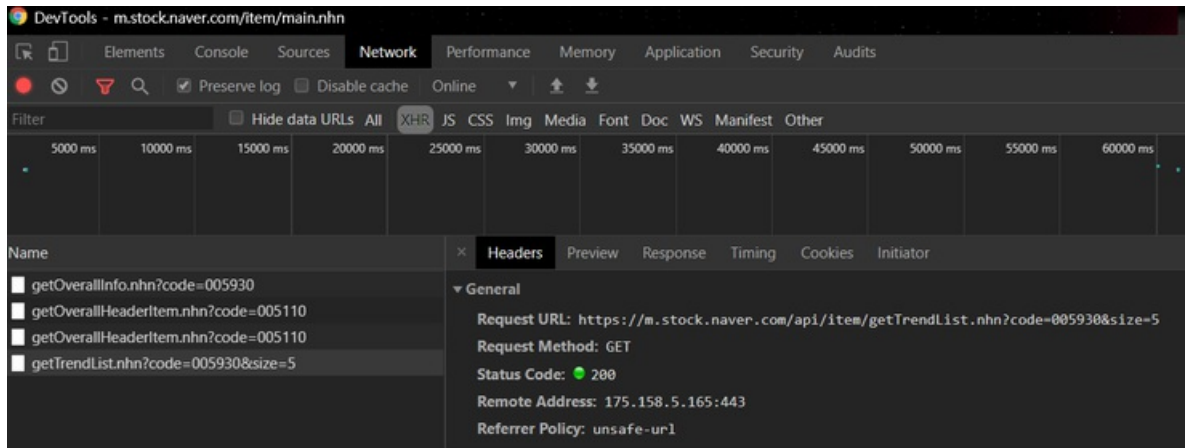
투자자별 매매동향					
날짜	span.foreign.stock_dn 66.95 x 16				
01.23.	-2,262,365	-57.16%	-1,906,336	+4,031,573	60,800
01.22.	+123,499	57.19%	+1,649,008	-1,731,848	62,300
01.21.	-905,210	-57.18%	-1,255,045	+2,003,000	60,800
01.20.	-213,817	-57.16%	+222,082	-88,000	60,800
01.17.	+976,840	-57.15%	-1,267,549	+164,000	60,800

크롬에서 검사 혹은 개발자도구로 보면 야속하게도 저렇게 값이 있는데 말이죠.

네이버 증권 PC화면 view에서는 비동기 처리를 안해도, mobile view 페이지(앞에 m.으로 시작되는 url) 에서는 해당 방식을 사용합니다. 그럼 굳이 mobile 페이지를 파싱 할 이유가 있느냐? 물으신다면 mobile 페이지에서 데이터를 가져오는 편이 속도면에서 훨씬 빠릅니다. 그리고 간혹 페이지 내에서 드러나지 않는 숨겨진 정보도 캐치할 수 있는 경우도 있고요.

그럼 이제 어떻게 해야하는지 알아보겠습니다.

## 2. JSON URL 가져오기



개발자 도구로 돌아갑니다. Network크 탭->XHR를 체크하고 불러들인 파일들을 확인해야 합니다.  
정확한 결과를 위해서 새로고침을 하고, 어떤 녀석들을 불러오는지 다시 확인을 하는 편이 좋은데요  
네이버 증권 페이지를 클릭하고 Ctrl+Shift+R을 눌러주면 캐시비우기를 하는 강력 새로고침을 하게 됩니다.

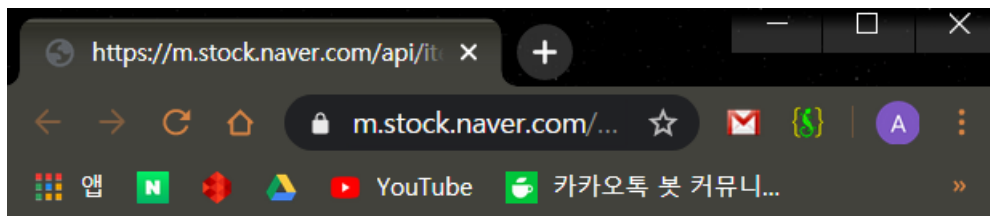
나온 결과들 중 파일들을 하나씩 클릭해봅니다.

Request URL 부분에 주소들이 나와있습니다.

이녀석들의 대부분은 Json 형식을 띄고 있는데, 각각 어떤 부분인지 URL 이름으로 추정해도 되고  
직접 하나하나 확인해보는 것도 재미있습니다.

저희는 투자자별 매매동향 정보를 알아보기 위함이므로, getTrendList가 적합해보입니다.

저 URL을 그대로 크롬으로 엽니다.



```
{
  "result": [
    {
      "change_val": -1500,
      "frgn_stock": 3412064282,
      "frgn_pure_buy_quant": -2262365,
      "frgn_hold_ratio": 57.15558738399944,
      "acc_quant": 14880396,
      "bizdate": "20200123",
      "rise_fall": "5",
      "itemcode": "005930",
      "listed_stock": 5969782,
      "close_val": 60800,
      "sosok": "01",
      "organ_pure_buy_quant": -1906336,
      "indi_pure_buy_quant": 4031573
    },
    {
      "change_val": 900,
      "frgn_stock": 3414326647,
      "acc_quant": 15339565,
      "bizdate": "20200122",
      "rise_fall": "2",
      "itemcode": "005930",
      "frgn_pure_buy_quant": 123499,
      "close_val": 62300,
      "sosok": "01",
      "frgn_hold_ratio": 57.19,
      "organ_pure_buy_quant": 1649008,
      "indi_pure_buy_quant": -1731848
    },
    {
      "change_val": -1000,
      "frgn_stock": 3413588080,
      "acc_quant": 11142693,
      "bizdate": "20200121",
      "rise_fall": "5",
      "itemcode": "005930",
      "frgn_pure_buy_quant": -905210,
      "close_val": 61400,
      "sosok": "01",
      "frgn_hold_ratio": 57.18,
      "organ_pure_buy_quant": -1255045,
      "indi_pure_buy_quant": 2003871
    },
    {
      "change_val": 1100,
      "frgn_stock": 3412169171,
      "acc_quant": 12528855,
      "bizdate": "20200120",
      "rise_fall": "2",
      "itemcode": "005930",
      "frgn_pure_buy_quant": -213817,
      "close_val": 62400,
      "sosok": "01",
      "frgn_hold_ratio": 57.16,
      "organ_pure_buy_quant": 222082,
      "indi_pure_buy_quant": -88665
    },
    {
      "change_val": 600,
      "frgn_stock": 3411538909,
      "acc_quant": 16025661,
      "bizdate": "20200117",
      "rise_fall": "2",
      "itemcode": "005930",
      "frgn_pure_buy_quant": 976840,
      "close_val": 61300,
      "sosok": "01",
      "frgn_hold_ratio": 57.15,
      "organ_pure_buy_quant": -1267549,
      "indi_pure_buy_quant": 164249
    }
  ],
  "resultCode": "success"
}
```

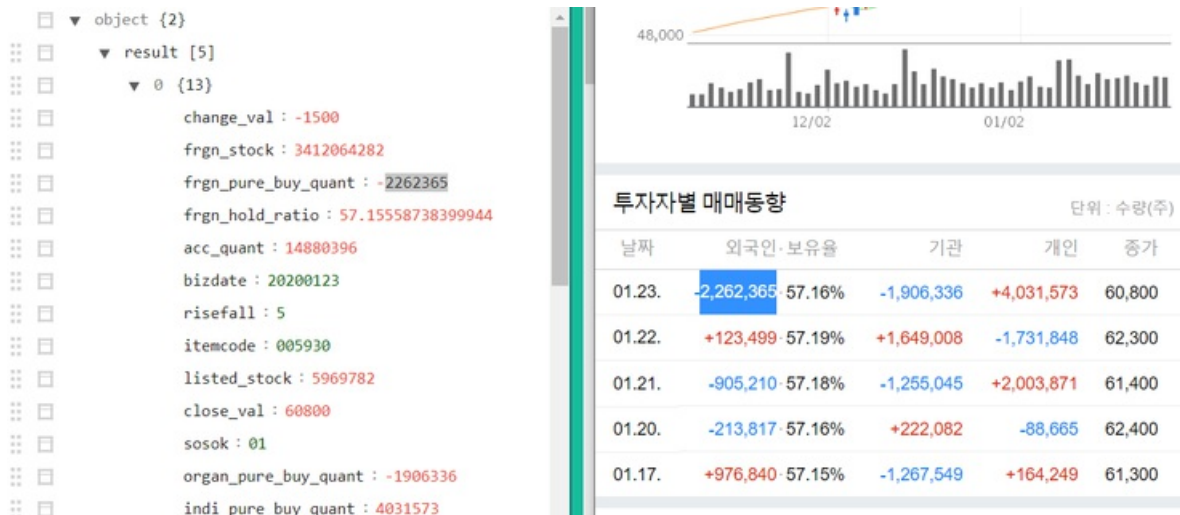
다른 API를 활용해보셨거나 JSON 형태가 익숙하신 분들이라면 쉽게 보실 수 있을겁니다.  
우리가 파싱하려는 정보가 저런 JSON 형식으로 전달되는데요

내가 원하는 값에만 정확히 접근할 수 있다면 오히려 HTML 파싱보다 더 효율적이고 직관적일 수 있습니다.

### 3. JSON 분석하여 활용하기

우선 저런 형태는 보기 어려우니 깔끔하게 보여주도록 합니다.

google에 json parser를 검색하고 원하는 페이지에 들어가서 저 내용(json)을 복사+붙여넣기 해줍니다.



좌측 화면이 파서를 이용한 화면입니다.

구조적으로 보기가 편리해졌죠? 증권 페이지에서 대조해보니 항목별로 잘 나와있는 것을 확인할 수 있습니다.

저는 1/23(최신 날짜) 외국인 동향을 알아보기 위해 frgn\_pure\_buy\_quant를 가져오려 합니다.

만약 저런 json의 구조에 대한 이해가 없으시다면 javascript 객체의 key:value 관련한 스터디를 하시길 바랍니다.

현재 저 json은 다음과 같은 구조를 가지고 있습니다.

우리가 원하는 값은 result라는 배열의 첫번째 ~result[0]~ 에 있는 frgn\_pure\_buy\_quant에 있습니다.

result[0] -> frgn\_pure\_buy\_quant

해당 값에 접근을 위해 다음과 같이 코드를 작성할 수 있습니다.

```
var url = https://m.stock.naver.com/api/item/getTrendList.nhn?code=005930&size=5;
var readJsoup = org.jsoup.Jsoup.connect(url).ignoreContentType(true).get().text();
// 해당 jsonURL을 Jsoup으로 파싱합니다.
//여기서 가장 중요한 것이 ignoreContentType(true)를 반드시 붙여줘야 하는 것 입니다.

var json = JSON.parse(readJsoup);
// json 형태 파싱법입니다. 인수로 전달받은 json 형태 문자열을 자바스크립트 객체로 변환

var newFrgn = json.result[0].frgn_pure_buy_quant;
// 객체로 변환했으니, 꺼먹을 수 있습니다.
```

제가 여기서 가장 해맸던 부분이..

.ignoreContentType(true)

이것을 몰라서 몇날몇일을 해맸는지 모릅니다.


꼭 붙여주셔야 합니다.





아미타님의 게시글 더보기 >


댓글 등록순 최신순 🔄


댓글알림 ☐


 **사로로**  
가끔 xhr말고 doc에 있을때도 있습니다. 네이버꺼가 그래요  
2020.01.26. 19:53 답글쓰기


 **아미타 작성자**  
그건 몰랐던 부분이네요. 감사합니다.  
2020.01.26. 19:54 답글쓰기


 **홍농**  
안 될 때마다 야매 헤딩으로 했었는데 확립된 정리글을 보니 마음이 편하군요..  
2020.01.26. 20:17 답글쓰기


 **아미타 작성자**  
예.. 저도 글로 남기니 다시 한 번 복습되네요 ㅎㅎ 감사합니다  
2020.01.26. 21:09 답글쓰기


 **Nub**  
꿀팁이네요.. 파싱 하시는 분들께는 도움 많이될듯  
2020.01.26. 20:58 답글쓰기

 **아미타 작성자**  
더 많은 노하우들이 공유되었으면 좋겠습니다. 좋게 봐주셔서 감사합니다.  
2020.01.26. 21:09 답글쓰기

 **현 흑**  
파싱이 안되는 페이지보단 파싱 주소를 찾는다는게 제목이 맞는거같아보이네요  
2020.01.26. 21:15 답글쓰기

 **아미타 작성자**  
의견 감사합니다. 처음 작성할 때 "왜 이페이지는 파싱이 안되지?" 하는 초보적인 관점으로 작성했는데 맞는 말씀인 것 같네요. 기존 제목에 추가해서 넣겠습니다.  
2020.01.26. 22:25 답글쓰기


 **현 흑**  
**아미타** 좋은글 감사합니다.  
2020.01.26. 23:10 답글쓰기


 **현 흑**  
**아미타** Network크 탭->XHR를 체크하고 불러들인 파일들을 확인해야 합니다.  
  
xhr doc보다는 해당부분에서 ctrl+f 해서 해당찾을 부분 검색해서 해당부분이 어느주소에서 호출되는지 찾는게 좀더 찾기 수월하더라구요.  
2020.01.26. 23:30 답글쓰기

 **kjh080713**



2020.01.27. 00:19 답글쓰기

 **delta**  
크롬 개발자도구에서 Network > Preview 탭을 이용하면 따로 Json 뷰어 없이도 편안하게 JSON 을 볼 수 있습니다.  
2020.01.28. 10:26 답글쓰기

 **아미타 작성자**



2020.01.28. 11:35 답글쓰기

Hibot

댓글을 남겨보세요



등록

글쓰기

답글

목록

▲ TOP

## '| 카카오톡 봇 강좌 |' 게시판 글

전체 [중급] 말머리 글

이 게시판 새글 구독하기 ☐

[꿀팁] (아마도) 컴터에서도 적용되는 ... 없는 전체보기! [9]	AlphaDo	2020.01.29.
온라인 카카오톡 트래커 #2-3 [5]	delta	2020.01.28.
[중급] 파싱 주소 찾기 (파싱이 안될 때; ajax; 비동기) 🙄 [13]	아미타	2020.01.26.
자바스크립트 in 연산자 [10]	이승기	2020.01.26.
채팅순위 용도 등의 객체 sorting [8]	이승기	2020.01.25.

1 2 3

전체보기

## 이 카페 인기글

요즘 나타나는 새로운 홍보 수법 중 하나가

Dark Tornado

♡0 💬7



[카카오톡 봇] 양산형 자동학습 봇 소스 공유

Dark Tornado

♡2 💬6

eval 질문

Kiri

♡0 💬5



5000분의 1을 얻은 사람

봇초보임니당

♡0 💬2

오픈이발 괴롭히기(메봇R)

ERROR

♡0 💬7



혹시 여기서 변수를 저장하는 부분이 어딘지 아시는분 계시나요...??

Milk2

♡0 💬4



틱택토 (Player vs Bot)

[ 치\*\*, 한\*\* ] 님 강제 탈퇴

AlphaDo  
♡ 1 💬 2

안녕하세요!

트위터봇  
♡ 0 💬 6

- 1
- 2
- 3
- 4
- 5