

파이썬 활용 사례

수업 목표

- 파이썬으로 웹 페이지를 **크롤링**하는 방법을 설명할 수 있다.
- 파이썬으로 **워드클라우드**를 만들 수 있다.
- 파이썬으로 **QR코드**를 만들고 URL로 연결할 수 있다.

1. 크롤링

1.1 크롤링이란?

1.2 웹 페이지 구조

1.3 웹페이지 크롤링하기

1.1 크롤링이란?

1.1 크롤링이란?

The screenshot shows the Naver English Dictionary interface. The search term 'crawl' is entered in the search bar, which has a green border. To the right of the search bar are a magnifying glass icon and a '고급 검색' (Advanced Search) button. Below the search bar, there is a navigation menu with tabs: '전체' (selected), '단어·속어', '뜻풀이', '예문', '유의어', and '영영사전'. To the right of the menu are three font size adjustment buttons labeled 'T', 'T', and 'T'. The main content area displays a list of related words and phrases: '연관검색어' followed by 'creep', 'crawling', 'gasp', 'knock over', 'gratitude', 'trip', 'stack', 'keep to oneself', 'turn out', 'walk along', 'craw', 'snap', 'end up', 'walk up to', 'sneak', 'hang out', 'jealous', 'crouch', 'encouragement', and 'creep'. Below this section is a '단어·속어' (Word/Slang) section with 79 items. A toggle switch for '글로벌 발음듣기' (Global Pronunciation) is turned on. The word 'crawl' is highlighted with a blue background and a red star icon. A list of definitions follows:

- 1. 동사 (엎드려) 기다
- 2. 동사 (곤충이) 기어가다
- 3. 명사 기어가기, 서행 (→pub crawl)
- 4. 명사 (수영의) 크롤법

1.1 크롤링이란?

크롤링(crawling)은 '기다'라는 뜻의 crawl의 명사형이다.

소프트웨어와 같은 무언가가 인터넷을 돌아다니며 정보를 수집해 오는 작업을 의미하고, 그러한 작업을 하는 소프트웨어를 **크롤러(crawler)**라고 한다.

월드와이드웹에서 웹페이지의 데이터를 '긁어' 오는 행위를 **스크레이핑(scraping)**이라고도 하는데 대체로 유사한 의미이다.

크롤링(crawling)은 웹 사이트(web site), 하이퍼링크(hyperlink), 데이터(data), 정보 자원을 자동화된 방법으로 수집, 분류, 저장하는 것을 의미함

웹 크롤링(web crawling) 또는 **데이터 크롤링(data crawling)**이라고도 함

**웹 상의 다양한 정보 자원을 자동화된 방법으로 수집해서
분류 및 저장하기 위함**

1.1 크롤링이란?

N | bts

통합 뉴스 이미지 VIEW 지식iN 인플루언서 동영상 쇼핑 어학사전 지도 ...



강원도민일보 PiCK | 6시간 전 | 네이버뉴스
새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적
 BTS)이 앨범 재킷 사진 촬영을 한 삼척 맹방해변을 찾는 국내·외 관광객들이 급증하고 있어 새로운 관광자원화 가능성을 높이고 있다. 삼척시에 따르면 일본인 관...
 삼척 맹방해변 'BTS 포토존' 새단장...日 관... 노컷뉴스 | 9시간 전 | 네이버뉴스
 삼척 맹방해변 BTS앨범 촬영지 새단장 강원일보 | 10시간 전 | 네이버뉴스
[관련뉴스 전체보기 >](#)



엑스포츠뉴스 | 11시간 전 | 네이버뉴스
방탄소년단(BTS) 진, 첫 솔로 앨범 발매 후 한터 주간차트 2관왕 ...
 그룹 방탄소년단(BTS) 멤버 진과 가수 원호(WONHO)가 한터차트 10월 5주 주간차트 1위에 등극했다. 지난 10월 31일 한터차트는 2022년 10월 5주 차 주간 월드차...
 방탄소년단(BTS) 진, 솔로 앨범 'The Astronaut'...한... 뉴스인사이드 | 14시간 전
 '한터차트' 스트레이키즈 2관왕, 블랙핑크·BTS 제쳤다 국제뉴스 | 4시간 전



스포츠조선 | 8시간 전 | 네이버뉴스
BTS 진 "슈퍼스타는 한정판 못 참아→돈 쓴 티 난 옷 좋아" 필수...

1.1 크롤링이란?

N bts

통합 뉴스 이미지 VIEW 지식iN 인플루언서 동영상 쇼핑 어학사전 지도 ...

강원도민일보 PiCK | 6시간 전 | 네이버뉴스

새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적

BTS)이 앨범 재킷 사진 촬영을 한 삼척 맹방해변을 찾는 국내·외 관광객들이 급증하고 있어 새로운 관광자원화 가능성을 높이고 있다. 삼척시에 따르면 일본인 관...

삼척 맹방해변 'BTS 포토존' 새단장...日 관... 노컷뉴스 | 9시간 전 | 네이버뉴스
삼척 맹방해변 BTS앨범 촬영지 새단장 강원일보 | 10시간 전 | 네이버뉴스

관련뉴스 전체보기 >

엑스포츠뉴스 | 11시간 전 | 네이버뉴스

방탄소년단(BTS) 진, 첫 솔로 앨범 발매 후 한터 주간차트 2관왕 ...

그룹 방탄소년단(BTS) 멤버 진과 가수 원호(WONHO)가 한터차트 10월 5주 주간차트 1위에 등극했다. 지난 10월 31일 한터차트는 2022년 10월 5주 차 주간 월드차...

방탄소년단(BTS) 진, 솔로 앨범 'The Astronaut'...한... 뉴스인사이드 | 14시간 전
'한터차트' 스트레이키즈 2관왕, 블랙핑크·BTS 제쳤다 국제뉴스 | 4시간 전

스포츠조선 | 8시간 전 | 네이버뉴스

BTS 진 "슈퍼스타는 한정판 못 참아→돈 쓴 티 난 옷 좋아" 필수...

1.2 웹 페이지 구조

1.2 웹 페이지 구조(html)

11

N bts

통합 뉴스 이미지 VIEW 지식iN 인플루언서 동영상 쇼핑 어학사전 지도 ...

강원도민일보 PICK | 6시간 전 | 네이버뉴스
새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적
BTS)이 앨범 재킷 사진 촬영을 한 삼척 맹방해변을 찾는 국내·외 관광객들이 급증하고 있어 새로운 관광자원화 가능성을 높이고 있다. 삼척시에 따르면 일본인 관...
삼척 맹방해변 'BTS 포토존' 새단장...日 관... 노컷뉴스 | 9시간 전 | 네이버뉴스
삼척 맹방해변 BTS 앨범 촬영지 새단장 강원일보 | 10시간 전 | 네이버뉴스
관련뉴스 전체보기 >

엑스포츠뉴스 | 11시간 전 | 네이버뉴스
방탄소년단(BTS) 진, 첫 솔로 앨범 발매 후 한터 주간차트 2관왕 ...
그룹 방탄소년단(BTS) 멤버 진과 가수 원호(WONHO)가 한터차트 10월 5주 주간차트 1위에 등극했다. 지난 10월 31일 한터차트는 2022년 10월 5주 차 주간 월드차...
방탄소년단(BTS) 진, 솔로 앨범 'The Astronaut'...한... 뉴스인사이드 | 14시간 전
'한터차트' 스트레이키즈 2관왕, 블랙핑크·BTS 제쳤다 국제뉴스 | 4시간 전

스포츠조선 | 8시간 전 | 네이버뉴스
BTS 진 "슈퍼스타는 한정판 못 참아→돈 쓴 티 난 옷 좋아" 필수...

뒤로 Alt+왼쪽 화살표
앞으로 Alt+오른쪽 화살표
새로고침 Ctrl+R
다른 이름으로 저장... Ctrl+S
인쇄 Ctrl+P
전송...
Google Lens로 이미지 검색

기기로 전송
이 페이지의 QR 코드 생성

한국어(으)로 번역

Google에서 이미지 설명 가져오기 ▶

페이지 소스 보기 Ctrl+U
검사

1.2 웹 페이지 구조(html)

뉴스

 강원도민일보 PiCK | 6시간 전 | 네이버뉴스

새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적

a.news_tit 468 x 25
Color #0C43B7
Font 16px -apple-system, BlinkMacSystemFont,...
Padding 1px 0px 0px
ACCESSIBILITY
Name 새단장한 삼척 BTS '버터비치', 일본인 ...
Role link
Keyboard-focusable

관련도순 • 최신



 엑스포즈뉴스 | 11시간 전 | 네이버뉴스

방탄소년단(BTS) 진, 첫 솔로 앨범 발매 후 한터 주간차트 2관왕 ...

그룹 방탄소년단(BTS) 멤버 진과 가수 원호(WONHO)가 한터차트 10월 5주 주간차트 1위에 등극했다. 지난 10월 31일 한터차트는 2022년 10월 5주 차 주간 월드차...

방탄소년단(BTS) 진, 솔로 앨범 'The Astronaut'...한... 뉴스인사이드 | 14시간 전
'한터차트' 스트레이키즈 2관왕, 블랙핑크·BTS 제쳤다 국제뉴스 | 4시간 전

 스포츠조선 | 8시간 전 | 네이버뉴스

BTS 진 "슈퍼스타는 한정판 못 참아→돈 쓴 티 난 옷 좋아" 필수...

GQ 유튜브 채널에서는 1일 '10 Things Jin (진) of BTS (방탄소년단) Can't Live With Out'이라는 제목의 영상이 게재됐다. 영상 속 진은 자신의 10가지 필수템들을 소...

BTS 진 "돈 쓴 티 나는 큰 로고 좋아" 구멍 ... TV리포트 | 5시간 전 | 네이버뉴스
"명품 바지에 구멍?"... BTS 진, 패션 신경 안쓰... 머니S | 5시간 전 | 네이버뉴스

DevTools is now available in Korean!

Always match Chrome's language [Switch DevTools to Korean](#) Don't show again

Elements Console Recorder Sources Network ↗ ⓘ

```

<div class="cm_top_wrap_sticky_custom_select_header">...</div>
<div class="cm_content_wrap">...</div>
<script>...</script>
</section>
<section class="sc_new sp_nnews_prs_nws_all">
  <div class="api_subject_bx">
    <div class="api_title_area">...</div>
    <div class="group_news">
      <ul class="list_news">
        <li class="bx" id="sp_nws_all1">
          <div class="news_wrap api_ani_send" style="flex">
            <div class="news_area">
              <div class="news_info">...</div>
              <a href="http://www.kado.net/news/articleView.html?idxno=1152179" class="news_tit" target="_blank" onclick="return goOtherCR(this, 'a=nws_all*h.tit&r=1&i=88221sgc_00000000000000000023848&g=654.0000023848&u='+urlencode(this.href));" title="새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적">...</a>
              <div class="news_dsc">...</div>
            </div>
            <a href="http://www.kado.net/news/articleView.html?idxno=1152179" class="dsc_thumb" target="_blank" onclick="return goOtherCR(this, 'a=nws_all*h.img&r=1&i=88221sgc_00000000000000000023848&g=654.0000023848&u='+urlencode(this.href));">...</a>
          </div>
        </li>
      </ul>
    </div>
  </div>
</section>

```

html body.wrap-new.api_animation div#wrap div#container div#content.pack_group

Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessibility

Filter :hov .cls + ↗ ⓘ

1.2 웹 페이지 구조(html)

13

```
<html>
<head>
<title>div example</title>

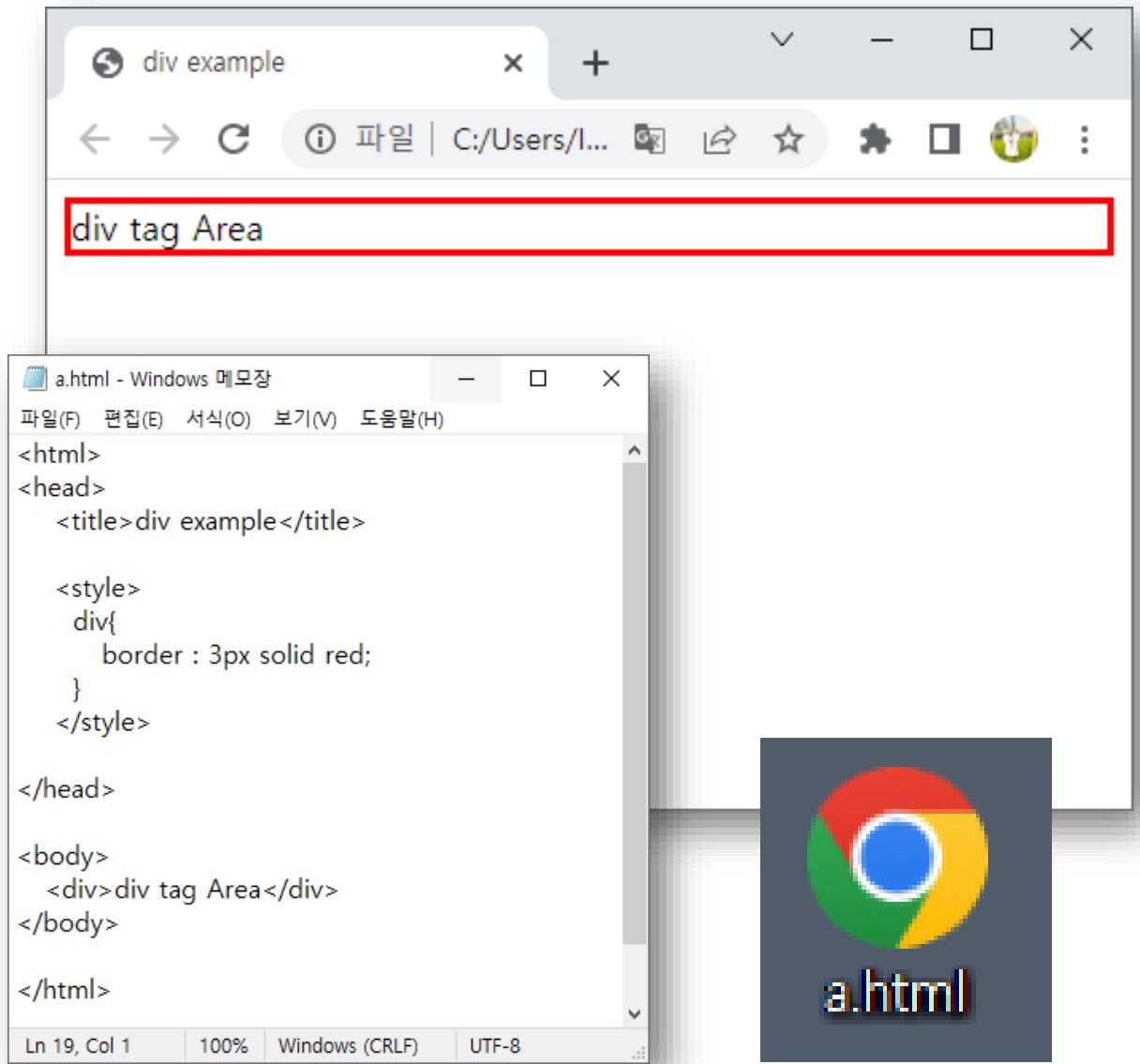
<style>
div{
    border : 3px solid red;
}
</style>

</head>

<body>
<div>div tag Area</div>
</body>

</html>
```

<출처> <https://yaegreena.tistory.com/52>



1.2 웹 페이지 구조(html)

14

```
<html>
<head>
<title>div example</title>

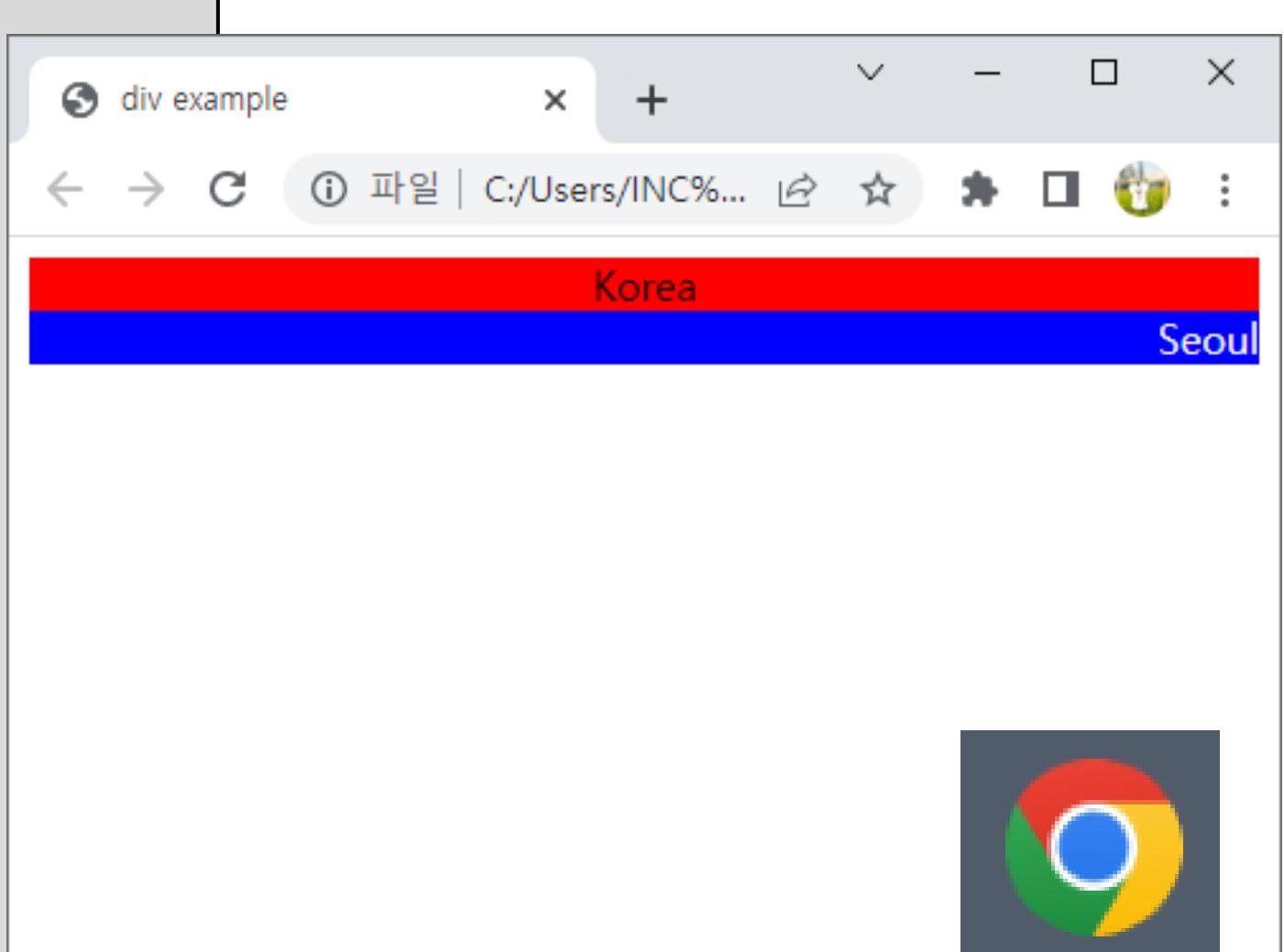
<style>
.nation {
background-color : red;
text-align: center;
}

.city {
background-color : blue;
text-align: right;
color:white;
}
</style>

</head>

<body>
<div class="nation">Korea</div>
<div class="city">Seoul</div>
</body>

</html>
```



<출처> <https://balmostory.tistory.com/100>



1.2 웹 페이지 구조(html)

15

```
<html>
<head>
<title>div example</title>

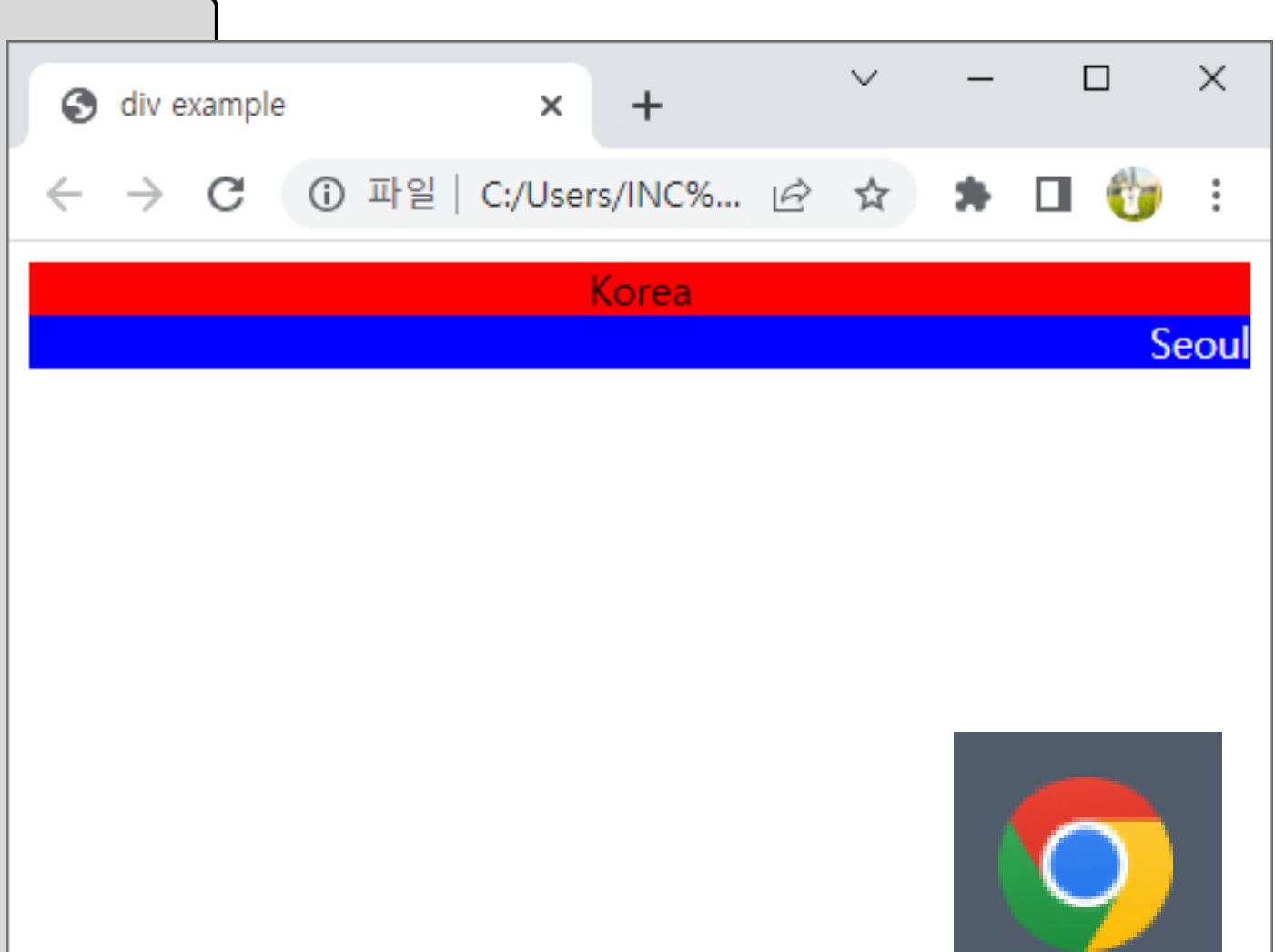
<style>
#nation {
background-color : red;
text-align: center;
}

#city {
background-color : blue;
text-align: right;
color:white;
}
</style>

</head>

<body>
<div id="nation">Korea</div>
<div id="city">Seoul</div>
</body>

</html>
```



1.2 웹 페이지 구조(html) - class vs id

16

```
<html>
<head>
<title>div example</title>

<style>
.nation {
background-color : red;
text-align: center;
}

.city {
background-color : blue;
text-align: right;
color:white;
}
</style>

</head>

<body>
<div class="nation">Korea</div>
<div class="city">Seoul</div>
</body>

</html>
```

.클래스 이름
(일반화된 내용 적용시)

```
<html>
<head>
<title>div example</title>
```

```
<style>
#nation {
background-color : red;
text-align: center;
}
```

```
#city {
background-color : blue;
text-align: right;
color:white;
}
</style>
```

```
</head>

<body>
<div id="nation">Korea</div>
<div id="city">Seoul</div>
</body>

</html>
```

#아이디 이름
(한 요소에 하나만 적용)

** ... : unodered list(순서가 없는 목록 태그)**

** ... : ordered list(순서가 있는 목록 태그)**

** ... : list item(목록에 해당하는 항목 태그)**

*.., ... 사이에 사용됨

1.2 웹 페이지 구조(html) - ul ol li

```
<html>
<head>
<title>div example</title>

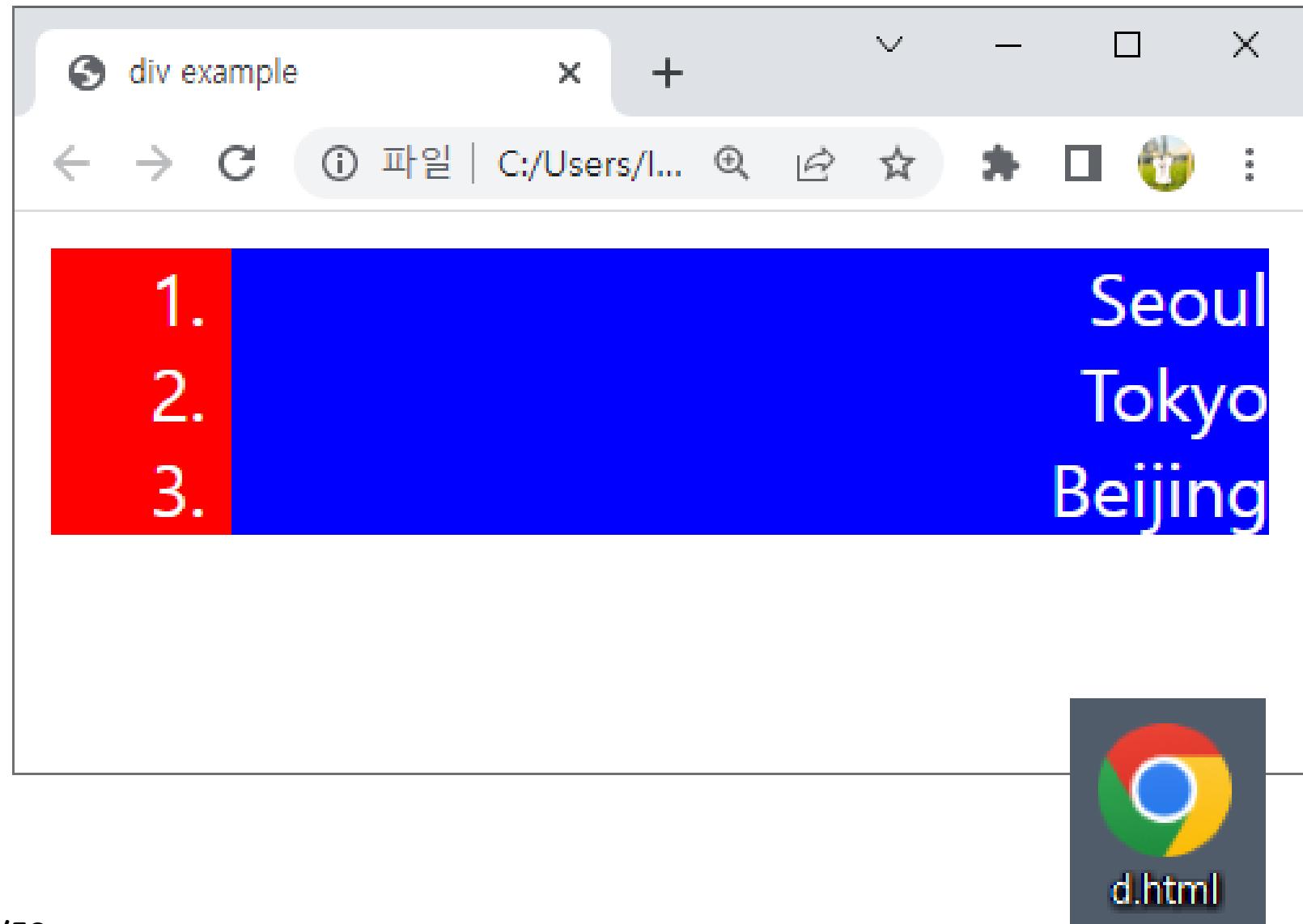
<style>
.nation {
background-color : red;
text-align: center;
}

.city {
background-color : blue;
text-align: right;
color:white;
}
</style>

</head>

<body>
<ol class = "nation">
<li class = "city">Seoul</li>
<li class = "city">Tokyo</li>
<li class = "city">Beijing</li>
</ol>
</body>

</html>
```



1.2 웹 페이지 구조(html)

19

1.2 웹 페이지 구조(html)

DevTools is now available in Korean! Always match Chrome's language Switch DevTools to Korean Don't show again

Elements Console Recorder Sources Network

```

<li class="bx" id="sp_nws_all1">
  <div class="news_wrap api_ani_send" style="flex">
    <div class="news_area">
      <div class="news_info">...</div>
      <a href="http://www.kado.net/news/articleView.html?idxno=1152179" class="news_tit" target="_blank" onclick="return goOtherCR(this, 'a=nws_all*h.tit&r=1&i=88221sgc_0000000000000000000023848&g=654.0000023848&u='+urlencode(this.href));" title="새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적">...</a>
      <div class="news_dsc">...</div>
    </div>
    <a href="http://www.kado.net/news/articleView.html?idxno=1152179" class="dsc_thumb" target="_blank" onclick="return goOtherCR(this, 'a=nws_all*h.img&r=1&i=88221sgc_0000000000000000000023848&g=654.0000023848&u='+urlencode(this.href));">...</a>
  </div>
</li>

```

Styles

:hov .cls

element.style {

search1...

html {

font-size : 10px;

line-height : 10px;

}

html[Attributes Style] {

-webkit-locale : "ko";

}

user agent...

html {

display : block;

}

1.2 웹 페이지 구조(html)

21

The screenshot shows a news website's homepage with several news items listed. The DevTools Elements tab is open, highlighting the `<div class="news_wrap api_ani_send">` element. This element is part of a larger news item structure. The right panel displays the full HTML code for the page, with the highlighted element's styling (flex) shown in the Styles panel.

News items visible on the page:

- 강원도민일보 PICK 6시간 전 | 네이버뉴스
새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적
BTS이 앨범 재킷 사진 촬영을 한 삼척 맹방해변을 찾는 국내외 관광객들이 급증하고 있어 새로운 관광자원화 가능성을 높이고 있다. 삼척시에 따르면 일본인 관...
- 삼척 맹방해변 'BTS 포토존' 새단장...日 관... 노컷뉴스 9시간 전 | 네이버뉴스
삼척 맹방해변 BTS앨범 촬영지 새단장 강원일보 10시간 전 | 네이버뉴스
- 관련뉴스 전체보기 >
- 엑스포즈뉴스 11시간 전 | 네이버뉴스
방탄소년단(BTS) 진, 첫 솔로 앨범 발매 후 한터 주간차트 2관왕 ...
그룹 방탄소년단(BTS) 멤버 진과 가수 원호(WONHO)가 한터차트 10월 5주 주간차트 1위에 등극했다. 지난 10월 31일 한터차트는 2022년 10월 5주 차 주간 월드차트...
- 방탄소년단(BTS) 진, 솔로 앨범 'The Astronaut'...한... 뉴스인사이드 14시간 전
'한터차트' 스트레이키즈 2관왕, 블랙핑크·BTS 제쳤다 국제뉴스 4시간 전
- 스포츠조선 8시간 전 | 네이버뉴스
BTS 진 "슈퍼스타는 한정판 못 참아→돈 쓴 티 난 옷 좋아" 필수...
GQ 유튜브 채널에서는 1일 '10 Things Jin (진) of BTS (방탄소년단) Can't Live With Out'이라는 제목의 영상이 게재됐다. 영상 속 진은 자신의 10가지 필수템들을 소...
- BTS 진 "돈 쓴 티 나는 큰 로고 좋아" 구멍 ... TV리포트 5시간 전 | 네이버뉴스
"명품 바지에 구멍?"... BTS 진, 패션 신경 안쓰... 머니S 5시간 전 | 네이버뉴스
- 관련뉴스 전체보기 >

DevTools Header:

- Always match Chrome's language
- Switch DevTools to Korean
- Don't show again

Elements Tab Headers:

- Elements
- Console
- Recorder
- Sources
- Network

Styles Panel:

```
:hover .cls {  
    element-style {  
        search1...  
        font-size : 10px;  
        line-height : 10px;  
    }  
    html {  
        font-size : 10px;  
        line-height : 10px;  
    }  
    html[AttributesStyle] {  
        -webkit-locale : "ko";  
    }  
    user agent...  
    html {  
        display : block;  
    }  
}
```

1.2 웹 페이지 구조(html)

22

The screenshot shows the Google Chrome DevTools interface with the 'Elements' tab selected. The left pane displays a news feed from Naver News, and the right pane shows the corresponding HTML code. A red box highlights the 'news_area' class in the CSS panel, which is defined as follows:

```
:hov .cls {  
element.s  
tyle {  
}  
search1_...  
html {  
font-  
size :  
10px;  
line-  
height :  
10px;  
}  
html[Attr  
ibutes  
Style] {  
-webkit-  
locale :  
"ko";  
}  
user age...  
html {  
display :  
block;  
}
```

The 'news_area' class is used to wrap news items, and its style is defined as a font size of 10px and a line height of 10px.

1.2 웹 페이지 구조(html)

23

뉴스

a.news_tit 468 × 25 보 PiCK | 6시간 전 | 네이버뉴스

새단장한 삼척 BTS '버터비치'; 일본인 관광객으로 북적

BTS)이 앨범 재킷 사진 촬영을 한 삼척 맹방해변을 찾는 국내·외 관광객들이 급증하고 있어 새로운 관광자원화 가능성을 높이고 있다. 삼척시에 따르면 일본인 관...

삼척 맹방해변 'BTS 포토존' 새단장...日 관... 노컷뉴스 | 9시간 전 | 네이버뉴스

삼척 맹방해변 BTS앨범 촬영지 새단장 강원일보 | 10시간 전 | 네이버뉴스

관련뉴스 전체보기 >

엑스포츠뉴스 | 11시간 전 | 네이버뉴스

방탄소년단(BTS) 진, 첫 솔로 앨범 발매 후 한터 주간차트 2관왕 ...

그룹 방탄소년단(BTS) 멤버 진과 가수 원호(WONHO)가 한터차트 10월 5주 주간차트 1위에 등극했다. 지난 10월 31일 한터차트는 2022년 10월 5주 차 주간 월드차...

방탄소년단(BTS) 진, 솔로 앨범 'The Astronaut'...한... 뉴스인사이드 | 14시간 전

'한터차트' 스트레이키즈 2관왕, 블랙핑크·BTS 제쳤다 국제뉴스 | 4시간 전

스포츠조선 | 8시간 전 | 네이버뉴스

BTS 진 "슈퍼스타는 한정판 못 참아→돈 쓴 티 난 옷 좋아" 필수...

GQ 유튜브 채널에서는 1일 '10 Things Jin (진) of BTS (방탄소년단) Can't Live With Out'이라는 제목의 영상이 게재됐다. 영상 속 진은 자신의 10가지 필수템들을 소...

BTS 진 "돈 쓴 티 나는 큰 로고 좋아" 구멍 ... TV리포트 | 5시간 전 | 네이버뉴스

"명품 바지에 구멍?"... BTS 진, 패션 신경 안쓰... 머니S | 5시간 전 | 네이버뉴스

관련뉴스 전체보기 >

1.3 웹페이지 크롤링하기

1.3 웹페이지 크롤링하기

N | bts

통합 뉴스 이미지 VIEW 지식IN 인플루언서 동영상 쇼핑 어학사전 지도 ...

강원도민일보 PICK 6시간 전 네이버뉴스
새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적
 BTS)이 앨범 재킷 사진 촬영을 한 삼척 맹방해변을 찾는 국내·외 관광객들이 급증하고 있어 새로운 관광자원화 가능성을 높이고 있다. 삼척시에 따르면 일본인 관...

삼척 맹방해변 'BTS 포토존' 새단장...日 관... 노컷뉴스 9시간 전 네이버뉴스
 삼척 맹방해변 BTS앨범 촬영지 새단장 강원일보 10시간 전 네이버뉴스

관련뉴스 전체보기 >

엑스포츠뉴스 11시간 전 네이버뉴스
방탄소년단(BTS) 진, 첫 솔로 앨범 발매 후 한터 주간차트 2관왕
 그룹 방탄소년단(BTS) 멤버 진과 가수 원호(WONHO)가 한터차트 10월 5주 주간차트 1위에 등극했다. 지난 10월 31일 한터차트는 2022년 10월 5주 차 주간 월드차...

방탄소년단(BTS) 진, 솔로 앨범 'The Astronaut'...한... 뉴스인사이드 14시간 전
 '한터차트' 스트레이키즈 2관왕, 블랙핑크·BTS 제쳤다 국제뉴스 4시간 전

스포츠조선 8시간 전 네이버뉴스
 BTS 진 "슈퍼스타는 한정판 못 참아→돈 쓴 티 난 옷 좋아" 필수...

CO crawling.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트 ✓ RAM 디스크 수정 가능 ^

[2] import requests
 from bs4 import BeautifulSoup

```
titles = []
search_word = 'bts'
url = f'https://search.naver.com/search.naver?where=nexearch&sm=top_hty&fbm=1&ie=utf8&query={search_word}'
#print(url)
req = requests.get(url)
html = req.text
#print(html)

soup = BeautifulSoup(html, 'html.parser')
search_result = soup.select_one('.list_news')
news_links = search_result.select('.bx > .news_wrap > .news_area > a')
#print(news_links)

for i in news_links:
    titles.append(i.get_text())
print(titles)
```

["새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적", "방탄소년단(BTS) 진, 첫 솔로 앨범 발매 후 한터 주간차트 2관왕"]

crawling

```
import requests  
from bs4 import BeautifulSoup
```

requests 모듈 : Python에서 HTTP 요청을 보내기 위해 사용

BeautifulSoup 라이브러리

- 루이스 캐럴의 『이상한 나라의 앤리스』에서 이름을 따옴
- 이야기 속 '모조 거북'이 이 노래를 부름
- BeautifulSoup는 잘못된 HTML을 수정하여 쉽게 탐색할 수 있는 XML 형식의 **파이썬 객체로 변환**하므로 골치 아픈 웹을 탐색할 때 유용함
- (지저분한 구조를 가진 웹문서를 두고 마구 뒤섞인 수프와 같다고 표현함)

아름다운 수프, 풍부한 녹색,
그릇에서 기다리거라!
누가 이 맛있는 것에 숙이지 않으리?
저녁 수프, 아름다운 수프!



```
#추출한 제목을 저장할 리스트
```

```
titles = []
```

```
#검색할 단어
```

```
search_word = 'bts'
```

```
#검색할 웹페이지 주소
```

```
url = f 'https://search.naver.com/search.naver?where=nexearch&sm=top_hty&fbm=1&ie=utf8&query={search_word}'
```

```
#검색어가 포함된 웹페이지 주소 확인
```

```
print(url)
```

1.3 웹페이지 크롤링하기

28

```
#requests 라이브러리를 통해 웹페이지 접근  
req = requests.get(url)  
html = req.text  
print(html)
```

```
[2]: import requests  
from bs4 import BeautifulSoup  
  
titles = []  
search_word = 'bts'  
url = f'https://search.naver.com/search.naver?where=nexearch&sm=top_hty&fbm=1&ie=utf8&query={search_word}'  
#print(url)  
req = requests.get(url)  
html = req.text  
print(html)
```

```
<!doctype html> <html lang="ko"> <head> <meta charset="utf-8"> <meta name="referrer" content="always"> <meta name="format-detection" content="telephone=no,address=no,email=no"> <meta name="viewport" content="width=device-width, initial-scale=1.0, maximum-scale=1.0, user-scalable=0" /> <span class="area_page" data-kgs-page-item></span> </script> <script>(function(){var startApplication=function(){var require>window.require.config({"context":'노출ID'로 변경해주세요.',"paths\":{\"@lapin-plus/flicking-page\":\"https://ssl.pstatic.net/lapinplus/flicking-page/dist/index.js\"}});</script>
```



```
<script> /* [AU] 인라인 스크립트 */ (function () { var CLASS_SECTION = "_sp_ntotal"; var CLASS_MARGIN = "type_margin"; var sectionEls = Array.prototype.slice.call(document.querySelectorAll("section")); var keywordEls = document.querySelectorAll("span.keyword"); var bxEls = document.querySelectorAll("span.bx"); for (var i = 0; i < sectionEls.length; i++) { var sectionEl = sectionEls[i]; var keywordEl = keywordEls[i]; var bxEl = bxEls[i]; if (sectionEl.classList.contains(CLASS_SECTION)) { sectionEl.classList.add(CLASS_MARGIN); } else { sectionEl.classList.remove(CLASS_MARGIN); } if (keywordEl.classList.contains(CLASS_MARGIN)) { keywordEl.classList.add(CLASS_SECTION); } else { keywordEl.classList.remove(CLASS_SECTION); } if (bxEl.classList.contains(CLASS_MARGIN)) { bxEl.classList.add(CLASS_SECTION); } else { bxEl.classList.remove(CLASS_SECTION); } } }); </script>
```

API

requests 라이브러리는 매우 직관적인 API를 제공하는데요. 어떤 방식(method)의 HTTP 요청을 하느냐에 따라서 해당하는 이름의 함수를 사용하면 됩니다.

- GET 방식: `requests.get()`
- POST 방식: `requests.post()`
- PUT 방식: `requests.put()`
- DELETE 방식: `requests.delete()`

1.3 웹페이지 크롤링하기

29

```
soup = BeautifulSoup(html, 'html.parser')
search_result = soup.select_one('.list_news')
news_links = search_result.select('.bx > .news_wrap > .news_area> .news_contents > .news_tit ')
print(news_links)
```

 crawling.ipynb 
파일 수정 보기 삽입 런타임 도구 도움말
+ 코드 + 텍스트

[2] import requests
from bs4 import BeautifulSoup

titles = []
search_word = 'bts'
url = f'https://search.naver.com/search.naver?where=nexearch&sm=top_hty&fbm=1&ie=utf8&query={search_word}'
#print(url)
req = requests.get(url)
html = req.text
#print(html)

soup = BeautifulSoup(html, 'html.parser')
search_result = soup.select_one('.list_news')
news_links = search_result.select('.bx > .news_wrap > .news_area > a')
print(news_links)

[...

```
▼<section class="sc_new sp_nnews _prs_nws_all">
  ▼<div class="api_subject_bx">
    ►<div class="api_title_area">...</div>
    ▼<div class="group_news">
      ▼<ul class="list_news">
        ▼<li class="bx" id="sp_nws_all1">
          ▼<div class="news_wrap api_ani_send" flex>
            ▼<div class="news_area">
              ►<div class="news_info">...</div>
              ►<a href="http://www.kado.net/news/articleview.php?nws_id=88221&tit=&r=1&i=88221sgc_0000000000003848&u='+urlencode(this.href);" title="새 일본인 관광객으로 북적">...</a> == $0
              ►<div class="news_dsc">...</div>
            ...
```

1.3 웹페이지 크롤링하기

30

```
for i in news_links:  
titles.append(i.get_text())  
print(titles)
```

for 변수 in 리스트:
반복할 코드

```
a = [38, 21, 53, 62, 19]  
for i in a :  
    print(i)
```

38
21
53
62
19

1.3 웹페이지 크롤링하기

31

```
for i in news_links:  
titles.append(i.get_text())  
print(titles)
```

for 변수 in 리스트:
반복할 코드

```
a = [38, 21, 53, 62, 19]  
for i in a :  
    print(i)
```

38
21
53
62
19

```
crawling.ipynb ☆  
파일 수정 보기 삽입 런타임 도구 도움말  
+ 코드 + 텍스트  
[2] import requests  
from bs4 import BeautifulSoup  
titles = []  
search_word = 'bts'  
url = f'https://search.naver.com/search.naver?where=nexearch&sm=top_hty&fbm=1&ie=utf8&query={search_#print(url)  
req = requests.get(url)  
html = req.text  
#print(html)  
  
soup = BeautifulSoup(html, 'html.parser')  
search_result = soup.select_one('.list_news')  
news_links = search_result.select('.bx > .news_wrap > .news_area > a')  
#print(news_links)  
  
for i in news_links:  
    titles.append(i.get_text())  
print(titles)  
["새단장한 삼척 BTS '버터비치', 일본인 관광객으로 북적", "BTS 제이홉 진 RM '개별 활동에 방탄소년단 보..."]
```

<출처> <https://www.youtube.com/watch?v=K50nD860U1A>

2. 워드 클라우드

- 2.1 워드 클라우드 만들기
- 2.2 워드 클라우드 만들기(파일기반)
- 2.3 워드 클라우드 만들기(파일기반/이미지 변경)
- 2.4 워드 클라우드 만들기(명사 빈도수 추출)
- 2.5 워드 클라우드 만들기(크롤링)

2. 워드 클라우드 만들기

태그 클라우드(영어: tag cloud) 또는 워드 클라우드(word cloud)는 태그들을 분석하여 중요도나 인기도 등을 고려하여 시각적으로 늘어놓아 표시하는 것이다.

시각적인 중요도를 강조를 위해 각 태그들은 그 중요도(혹은 인기도)에 따라 글자의 색상이나 굵기 등 형태가 변한다.



A screenshot of a Google search results page. The search bar at the top contains the query "워드클라우드". Below the search bar are several navigation links: 전체 (selected), 이미지, 뉴스, 동영상, 지도, 더보기, and 도구. The main search result is a link to "http://wordcloud.kr" with the title "워드클라우드 생성기". The snippet below the title describes the site as a word cloud generator that creates word clouds from Korean text, mentioning features like 한글, 단어구름, 글자구름, 구름글자, 태그클라우드, and 랜덤이미지.

2.1 워드 클라우드 만들기 - 한글 폰트 설치하기

35

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** wordcloud.ipynb
- File Menu:** 파일, 수정, 보기, 삽입, 런타임, 도구, 도움말, 모든 변경사항이 저장됨
- Toolbar:** 대화, 공유, 설정 가능
- File Explorer:** 파일, sample_data, NanumGothicExtraBold.ttf
- Code Cell:** [1] from wordcloud import WordCloud
import matplotlib.pyplot as plt

#<https://www.youtube.com/watch?v=E5ABFfZFeoQ>
a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때 기분이 좋아요'

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white")
wordcloud = wordcloud.generate(a)

plt.imshow(wordcloud)
plt.axis('off')
plt.show()
- Message Bar:** 업로드 완료된 후 코드 실행 (Yellow Callout)
- Disk Usage:** 디스크 85.16 GB 사용 가능
- Execution Status:** 0초 오후 7:22에 완료됨

<출처> <https://www.youtube.com/watch?v=E5ABFfZFeoQ>

2.1 워드 클라우드 만들기 - 한글 폰트 설치하기

36

The screenshot shows a Jupyter Notebook interface with the following details:

- File Explorer:** On the left, there is a file tree. A file named "NanumGothicExtraBold.ttf" is selected and highlighted with a red border.
- Code Cell:** The main area contains Python code for generating a word cloud.

```
[1]: from wordcloud import WordCloud
import matplotlib.pyplot as plt

#https://www.youtube.com/watch?v=E5ABFfZFeoQ
a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때 기분이 좋아요'

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white")
wordcloud = wordcloud.generate(a)

plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```
- Output:** Below the code cell, the status bar indicates "0초 오후 7:22에 완료됨" (0 seconds completed at 7:22 PM).
- Annotations:** A yellow callout bubble points to the line "font_path='NanumGothicExtraBold.ttf'" with the text "업로드 된 파일명과 동일" (Same as uploaded file name).

<출처> <https://www.youtube.com/watch?v=E5ABFfZFeoQ>

2.1 워드 클라우드 만들기

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** 14_1_1_wordcloud.ipynb
- File Menu:** 파일, 수정, 보기, 삽입, 런타임, 도구, 도움말
- Code Cell:** [1] from wordcloud import WordCloud
import matplotlib.pyplot as plt
- Text Cell:** #<https://www.youtube.com/watch?v=E5ABFfZFeoQ>
a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때 기분이 좋아요'

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white")
wordcloud = wordcloud.generate(a)

plt.imshow(wordcloud)
plt.axis('off')
plt.show()
- Output:** A generated word cloud image where the most frequent words are "수학을" (math), "저는" (I), and "좋아합니다" (like).
- Bottom Status Bar:** 0초 오후 7:57에 완료됨

A yellow callout bubble points to the code cell with the text: "파이썬에서 제공하는 word cloud 모듈을 사용하여 word cloud를 만들 수 있음".

<출처> <https://www.youtube.com/watch?v=E5ABFfZFeoQ>

2.1 워드 클라우드 만들기

38



<출처> <https://www.youtube.com/watch?v=E5ABFfZFeoQ>

2.1 워드 클라우드 만들기

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** 14_1_1_wordcloud.ipynb
- File Menu:** 파일, 수정, 보기, 삽입, 런타임, 도구, 도움말
- Toolbar:** + 코드, + 텍스트, RAM, 디스크, 수정 가능
- Code Cell 1:** [1] from wordcloud import WordCloud
import matplotlib.pyplot as plt
- Text Cell 1:** #<https://www.youtube.com/watch?v=E5ABFfZFeoQ>
a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때 기분이 좋아요'

#wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white")
wordcloud = wordcloud.generate(a)

plt.imshow(wordcloud)
plt.axis('off')
plt.show()
- Output:** A word cloud visualization where the most frequent words are "수학을" (math), "저는" (I), and "좋아합니다" (like).
- Bottom Status Bar:** 0초 오후 7:57에 완료됨

A yellow callout bubble highlights the text "분석할 문장을 변수 a에 저장".

<출처> <https://www.youtube.com/watch?v=E5ABFfZFeoQ>

2.1 워드 클라우드 만들기

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** 14_1_1_wordcloud.ipynb
- File Menu:** 파일, 수정, 보기, 삽입, 런타임, 도구, 도움말
- Toolbar:** + 코드, + 텍스트, RAM, 디스크, 수정 가능
- Code Cell:** [1] from wordcloud import WordCloud
import matplotlib.pyplot as plt

#<https://www.youtube.com/watch?v=E5ABFfZFeoQ>
a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때 기분이 좋아요'

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white")
wordcloud = wordcloud.generate(a)

plt.imshow(wordcloud)
plt.axis('off')
plt.show()
- Output Cell:** A word cloud visualization where the words "좋아합니다", "참", "저는", "풀", "수학", and "문제를" are rendered in green and purple, set against a white background.
- Text Overlay:** 워드 클라우드 이미지 크기&배경색 (Word Cloud Image Size & Background Color)
- Bottom Status Bar:** 0초 오후 7:57에 완료됨

<출처> <https://www.youtube.com/watch?v=E5ABFfZFeoQ>

2.1 워드 클라우드 만들기

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** 14_1_1_wordcloud.ipynb
- File Menu:** 파일, 수정, 보기, 삽입, 런타임, 도구, 도움말
- Toolbar:** 파일, 디스크, 수정 가능
- Code Cell:** [1] from wordcloud import WordCloud
import matplotlib.pyplot as plt

#<https://www.youtube.com/watch?v=E5ABFfZFeoQ>
a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때 기분이 좋아요'

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white")
wordcloud = wordcloud.generate(a)

plt.imshow(wordcloud)
plt.axis('off')
plt.show()
- Output Cell:** A word cloud visualization where the most frequent words are "좋아합니다", "참", "저는", "수학", and "풀". Smaller words like "기분이", "좋아요", and "문제를" are also visible.
- Bottom Status Bar:** 0초 오후 7:57에 완료됨

<출처> <https://www.youtube.com/watch?v=E5ABFfZFeoQ>

2.1 워드 클라우드 만들기

The screenshot shows a Jupyter Notebook interface with the following details:

- Title:** 14_1_1_wordcloud.ipynb
- Code Cell Content:**

```
[1] from wordcloud import WordCloud
import matplotlib.pyplot as plt

#https://www.youtube.com/watch?v=E5ABFfZFeoQ
a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때 기분이 좋아요'

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white")
wordcloud = wordcloud.generate(a)

plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```
- Output:** A word cloud visualization where the words "저는", "수학을", and "좋아합니다" are the largest, all in green. Smaller words include "참", "좋아요", "풀", "때", "문제를", and "기분이".
- Annotations:** A yellow callout box highlights the line `plt.axis('off')` with the text "축(axis) 표시 : off".
- File Explorer:** Shows a folder structure with `sample_data` and `NanumGothicExtraBold.ttf`.
- System Status:** RAM usage is at 0%, Disk usage is at 85.18 GB.

<출처> <https://www.youtube.com/watch?v=E5ABFfZFeoQ>

2.1 워드 클라우드 만들기

CO 14_1_1_wordcloud.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

댓글 공유 설정

파일

코드 텍스트

RAM 디스크 수정 가능

[1] from wordcloud import WordCloud
import matplotlib.pyplot as plt

#<https://www.youtube.com/watch?v=E5ABFfZFeoQ>
a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때 기분이 좋아요'

#wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white")
wordcloud = wordcloud.generate(a)

plt.imshow(wordcloud)
plt.axis('on') **축(axis) 표시 : on**
plt.show()

문제를 수학 참 좋아요 저는 수학을 좋아합니다 기분이 좋았습니다

디스크 85.18 GB 사용 가능

0초 오후 7:56에 완료됨

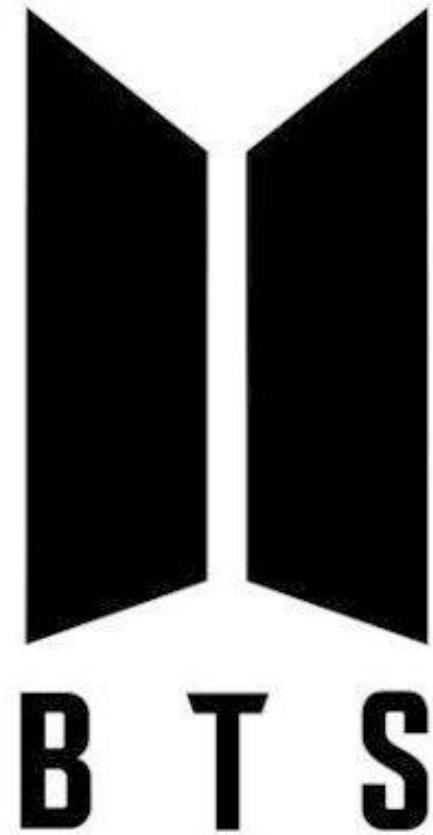
<출처> <https://www.youtube.com/watch?v=E5ABFfZFeoQ>

2.2 워드 클라우드 만들기 (파일기반)

2.2 워드 클라우드 만들기(파일기반)

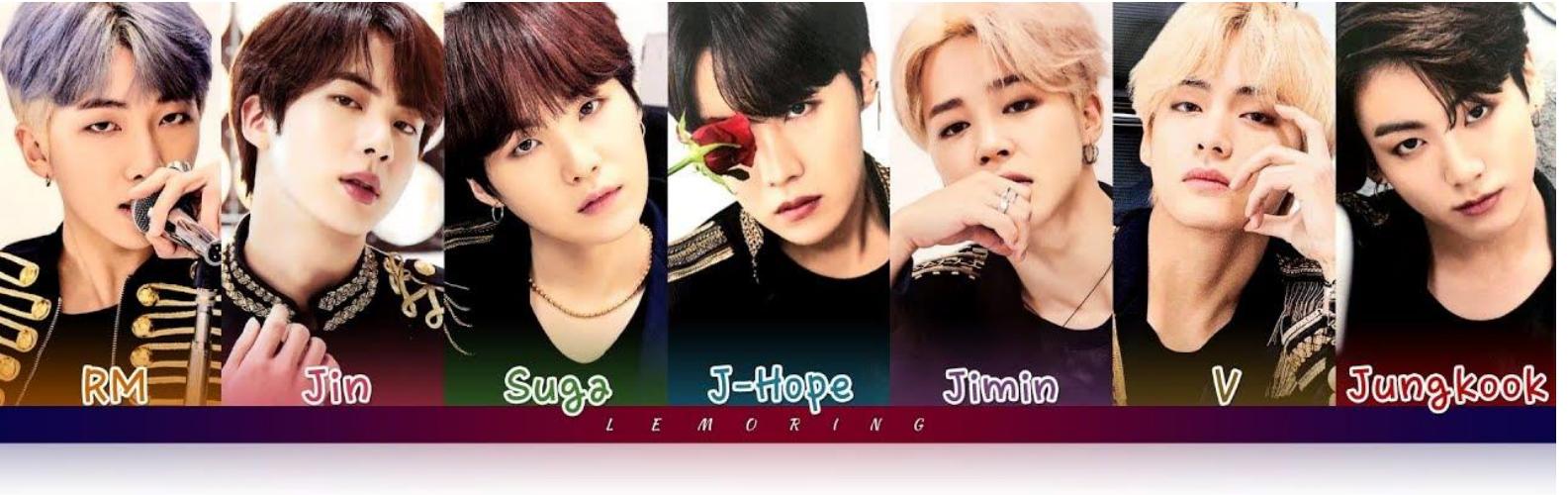
전반적으로 작곡/작사에서 BTS는 70%이상 참여하고 있고, GOT7는 20~30%, EXO는 10% 미만 참여하는 것을 알 수 있습니다. 전체 신곡 기준으로 보더라도, 남자 아이돌까지 비교했을 때 BTS 참여도가 비교적 높았습니다.

<출처><https://observablehq.com/@shwldus3/02-2-bts>



2.2 워드 클라우드 만들기(파일기반)

46



BLOOD SWEAT & TEARS

BTS (방탄소년단)

“피, 땀, 눈물”에서 가장 많이 나오는 단어는?

2.2 워드 클라우드 만들기(파일기반)

내 피 땀 눈물 내 마지막 춤을 다 가져가 가
내 피 땀 눈물 내 차가운 숨을 다 가져가 가
내 피 땀 눈물

내 피 땀 눈물도
내 몸 마음 영혼도
너의 것인 걸 잘 알고 있어
이건 나를 벌받게 할 주문
Peaches and cream
Sweeter than sweet
Chocolate cheeks
and chocolate wings
But 너의 날개는 악마의 것
너의 그 sweet 앞엔 bitter bitter
Kiss me 아파도 돼 어서 날 조여줘
더 이상 아플 수도 없게
Baby 취해도 돼 이제 널 들이켜
목 깊숙이 너란 위스키

내 피 땀 눈물 내 마지막 춤을 다 가져가 가
내 피 땀 눈물 내 차가운 숨을 다 가져가 가

원해 많이 많이 많이
원해 많이 많이 많이 많이
원해 많이 많이 많이
원해 많이 많이 많이 많이

아파도 돼 날 묶어줘 내가 도망칠 수 없게
꽉 쥐고 날 흔들어줘 내가 정신 못 차리게
Kiss me on the lips lips 둘만의 비밀
너란 감옥에 중독돼 깊이
니가 아닌 다른 사람 섬기지 못해
알면서도 삼켜버린 독이 든 성배

내 피 땀 눈물 내 마지막 춤을 다 가져가 가
내 피 땀 눈물 내 차가운 숨을 다 가져가 가

원해 많이 많이 많이
원해 많이 많이 많이 많이
원해 많이 많이 많이
원해 많이 많이 많이 많이

나를 부드럽게 죽여줘
너의 손길로 눈 감겨줘
어차피 거부할 수조차 없어
더는 도망갈 수조차 없어
니가 너무 달콤해 너무 달콤해
너무 달콤해서

내 피 땀 눈물
내 피 땀 눈물

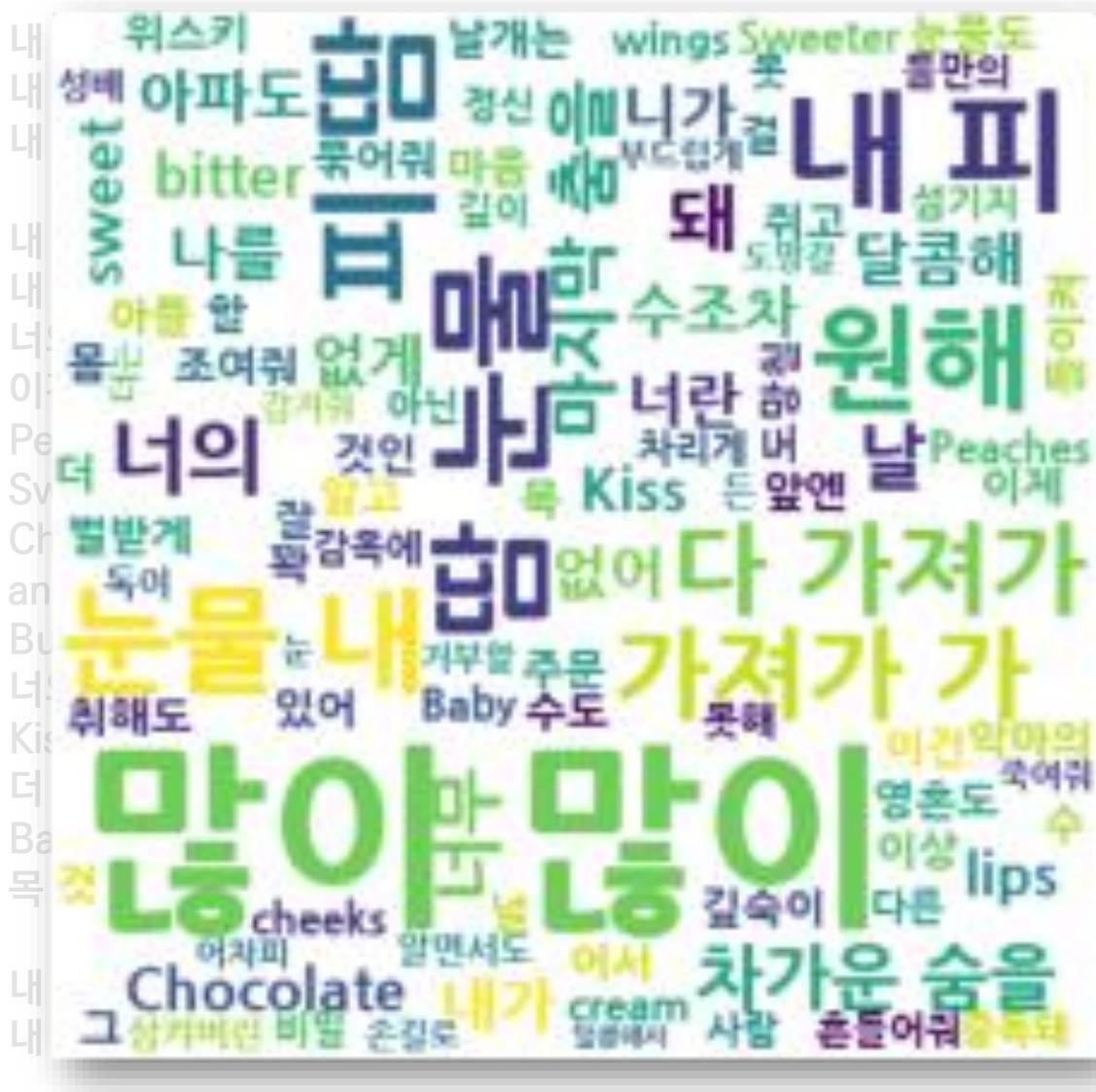


BLOOD SWEAT & TEARS

BTS (방탄소년단)

2.2 워드 클라우드 만들기(파일기반)

48



많이 많이
많이 많이 많이 많이
많이 많이
많이 많이
많이 많이

어줘 내가 도망칠 수 없게
어줘 내가 정신 못 차리게
the lips lips 둘만의 비밀
독돼 짖이
사람 심기 못해
내린 독이 든 8배
“피, 땀,

마지막 춤을 다 가져가
차가운 숨을 다 가져가

많이 많이
많이 많이 많이
많이 많이
많이 많이
많이 많이 많이

나를 부드럽게 죽여줘
너의 손길로 눈 감겨줘
어차피 거부할 수조차 없어
더는 도망갈 수조차 없어
니가 너무 달콤해 너무 달콤해
너무 달콤해서

“ㅍ, 땀, 눈물” < “땀이”

2.2 워드 클라우드 만들기(파일기반)

49

BTS PLAYLIST



01. Yet To Come	16. Save Me
02. Boy With Luv (Ft. Halsey)	17. I'm Fine
03. Permission To Dance	18. So What
04. Butter	19. UGH!
05. Dynamite	20. Not Today
06. Spring Day	21. Dionysus
07. DNA	22. Anpanman
08. FAKE LOVE	23. Go Go
09. MIC Drop (Remix)	24. Airplane pt.2
10. IDOL	25. War of Hormone
11. Blood Sweat & Tears	26. Boy In Luv
12. FIRE	27. No More Dream
13. On	28. Danger
14. DOPE	29. RUN
15. Run BTS	30. Life Goes On

2.2 워드 클라우드 만들기(파일기반)

50

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** CO 14_1_2_wordcloud_textFile.ipynb ☆
- File Menu:** 파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨
- Toolbar:** 파일, 디스크, 수정 가능, 등록된 파일, 공유, 설정, 프로필
- File Explorer:** 파일 목록 (sample_data, NanumGothicExtraBold.ttf, bts.txt, bts.txt)
- Code Cell 0초:** [1]

```
from wordcloud import WordCloud  
import matplotlib.pyplot as plt
```
- Code Cell 1초:** [2]

```
#a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때'  
with open('bts.txt', 'r', encoding='utf-8') as f:  
    a = f.read()
```

A yellow callout box points to this line with the text "파일 불러오기".

```
wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white")  
wordcloud = wordcloud.generate(a)  
  
plt.imshow(wordcloud)  
plt.axis('off')  
plt.show()
```

2.3 워드 클라우드 만들기 (파일기반/이미지 변경)

2.3 워드 클라우드 만들기(파일기반/이미지 변경)

52

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** 14_1_3_wordcloud_textFile_img.ipynb
- File Explorer:** Shows a directory structure with sample_data, NanumGothicExtraBold.ttf, bts.txt, and heart.png.
- Code Cell:** Contains Python code for generating a word cloud.

```
[5] from wordcloud import WordCloud
import matplotlib.pyplot as plt
from PIL import Image
import numpy as np

#a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때'
with open('bts.txt', 'r', encoding='utf-8') as f:
    a = f.read()

img = Image.open('heart.png') #추가
img_array = np.array(img) #추가

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white", mask=img_array)
wordcloud = wordcloud.generate(a)

plt.figure() #추가
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```
- Callout Box:** A yellow callout box highlights the line `from PIL import Image` with the text "이미지 분석 및 처리를 지원하는 라이브러리 (Python Imaging Library : PIL)".
- Toolbar:** Includes buttons for file operations, code execution, and notebook management.

2.3 워드 클라우드 만들기(파일기반/이미지 변경)

53

The screenshot shows a Google Colab notebook titled "14_1_3_wordcloud_textFile_img.ipynb". The left sidebar displays a file tree with "sample_data", "NanumGothicExtraBold.ttf", "bts.txt", and "heart.png". The main area contains the following Python code:

```
[5] from wordcloud import WordCloud
import matplotlib.pyplot as plt
from PIL import Image
import numpy as np

#a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때'
with open('bts.txt', 'r', encoding='utf-8') as f:
    a = f.read()

img = Image.open('heart.png') #추가
img_array = np.array(img)    #추가

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white", mask=img_array)
wordcloud = wordcloud.generate(a)

plt.figure() #추가
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```

A yellow callout box highlights the imports for Pandas and Matplotlib, with the text: "벡터 및 행렬 연산을 위한 라이브러리 (Pandas, Matplotlib 등의 기반이 됨)".

2.3 워드 클라우드 만들기(파일기반/이미지 변경)

54

The screenshot shows a Jupyter Notebook interface with the following details:

- File Explorer:** On the left, under the folder `{x}`, there are files: `..`, `sample_data`, `NanumGothicExtraBold.ttf`, `bts.txt`, and `heart.png`. The `heart.png` file is highlighted with a red rectangle.
- Code Cell:** The main area contains a code cell numbered [5] with the following Python code:

```
[5] from wordcloud import WordCloud
import matplotlib.pyplot as plt
from PIL import Image
import numpy as np

#a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때'
with open('bts.txt', 'r', encoding='utf-8') as f:
    a = f.read()

img = Image.open('heart.png') #추가
img_array = np.array(img) #추가

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white", mask=img_array)
wordcloud = wordcloud.generate(a)

plt.figure() #추가
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```
- Annotations:** A yellow callout box with the text "이미지 추가" (Image added) points to the line `img = Image.open('heart.png')`.
- Toolbar:** At the top right, there are buttons for RAM, Disk, and Edit mode.

2.3 워드 클라우드 만들기(파일기반/이미지 변경)

55

The screenshot shows a Jupyter Notebook interface with the following details:

- File Explorer:** On the left, it shows a directory structure with a folder named "sample_data" containing "NanumGothicExtraBold.ttf", "bts.txt", and "heart.png".
- Code Cell:** The main area contains a code cell with the following Python code:

```
[5] from wordcloud import WordCloud
import matplotlib.pyplot as plt
from PIL import Image
import numpy as np

#a = '저는 수학을 참 좋아합니다. 수학 문제를 풀 때'
#with open('bts.txt', 'r', encoding='utf-8') as f:
#    a = f.read()

img = Image.open('heart.png') #추가
img_array = np.array(img) #추가

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf")
wordcloud = wordcloud.generate(a)

plt.figure() #추가
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```
- Output:** To the right of the code cell, a word cloud visualization is displayed. It features large, bold Korean words like "다 가져가", "만이 많아", and "나는 땀 눈물 피 땀 눈물 피" in green and blue, overlaid on a background image of a heart.

2.3 워드 클라우드 만들기(파일기반/이미지 변경)

56

눈물	10
마지막	3
마음	1
영혼	1
이건	1
주문	1
날개	1
악마	1
어서	1
이상	1
수도	1
취해	1
이제	1

이제	1
위스키	1
정신	1
비밀	1
감옥	1
중독	1
깊이	1
다른	1
사람	1
성배	1
손길	1
어차피	1
거부	1



2.4 워드 클라우드 만들기 (명사 빈도수 추출)

2.4 워드 클라우드 만들기(명사 빈도수 추출)

사랑	67	여기	20
모든	57	누구	20
방탄소년단	54	오늘	19
작곡	54	위로	19
작사	53	시소	19
우리	53	어디	18
사람	47	소리	18
우린	43	생각	18
세상	37	노래	18
이제	33	분노	18
다시	30	겨울	17
방시혁	30	누가	17
탕진	28	모두	17
욱해	28	조심	17
위해	25	하나	16
지금	24	하루	16
심장	24	바람	16
전부	23	그때	16
순간	23	우연이	15
계속	23	가장	14
운명	23	다른	14
고민	22	진짜	14
그냥	21	자꾸	13

2.4 워드 클라우드 만들기(명사 빈도수 추출)

59

The screenshot shows a Google Colab notebook titled "14_1_4_text_frequency_korean.ipynb". The code cell [61] reads a file named "bts_all.txt" and stores its content in the variable "lines". The code cell [62] installs the "konlpy" package via pip. The output of this cell shows the package is already satisfied. The code cell [63] imports the Twitter class from konlpy.tag and prints the nouns found in the text. A warning message is displayed about the "Twitter" class being renamed to "Okt".

```
[61] #https://dalulu.tistory.com/108
f = open("bts_all.txt", "r")

lines = f.read()

[62] !pip install konlpy

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: konlpy in /usr/local/lib/python3.7/dist-packages (0.6.0)
Requirement already satisfied: JPype1>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from konlpy) (1.4.1)
Requirement already satisfied: lxml>=4.1.0 in /usr/local/lib/python3.7/dist-packages (from konlpy) (4.9.1)
Requirement already satisfied: numpy>=1.6 in /usr/local/lib/python3.7/dist-packages (from konlpy) (1.21.6)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packages (from JPype1>=0.7.0>konlpy) (4.1.1)
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from JPype1>=0.7.0>konlpy) (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging->JPype1>=0.7.0>konlpy) (3.0.9)

[63] from konlpy.tag import Twitter
nlp = Twitter()
nouns = nlp.nouns(lines)
print(nouns)

/usr/local/lib/python3.7/dist-packages/konlpy/tag/_okt.py:17: UserWarning: "Twitter" has changed to "Okt" since KoNLPy v0.4.5.
  warn("Twitter" has changed to "Okt" since KoNLPy v0.4.5.)
['방탄소년단', '작사', '작곡', '날', '세상', '예고', '하나', '봄', '기다림', '눈치', '발자국', '거리', '여기', '나', '흔자', '시간', '말', '오늘', '비', '']
```

2.4 워드 클라우드 만들기(명사 빈도수 추출)

60

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** 14_1_4_text_frequency_korean.ipynb
- File Menu:** 파일, 수정, 보기, 삽입, 런타임, 도구, 도움말, 모든 변경사항이 저장됨
- Toolbar:** 파일, 코드, 텍스트, RAM, 디스크, 수정 가능
- File Explorer:** sample_data, NanumGothicExtraBold.ttf, bts_all.txt, heart.png
- Code Cell 64:** Imports Counter from collections and initializes count = Counter(nouns). Creates lists tag_count and tags. Iterates over the most common nouns (100) and creates a dictionary dics with 'tag' and 'count'. If the length of tag_count is 2 or less and the length of tags is 49 or less, it appends dics to both lists.
- Code Cell 65:** Prints each tag and its count using formatted strings. The output shows the top 15 nouns and their counts.
- Output:** A table of the top 15 nouns and their counts.

단어	빈도수
사랑	67
모든	57
방탄소년단	54
작곡	54
작사	53
우리	53
사람	47
우린	43
세상	37
이제	33
다시	30
방시혁	30

2.4 워드 클라우드 만들기(명사 빈도수 추출)

61

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** 14_1_4_text_frequency_korean.ipynb
- Toolbar:** 파일, 수정, 보기, 삽입, 런타임, 도구, 도움말
- File Explorer:** 파일, sample_data, NanumGothicExtraBold.ttf, bts_all.txt, heart.png
- Code Cells:**
 - [66] word_list = []
for tag in tag_count:
 b = tag['tag'] + ''
 word_list.append(b*tag['count'])

#print(word_list)
a = ' '.join(s for s in word_list)
print(a)

The output of this cell is a long string of the word "사랑" repeated many times.
 - [67] from wordcloud import WordCloud
import matplotlib.pyplot as plt
from PIL import Image
import numpy as np
 - [69] img = Image.open('heart.png') #추가
img_array = np.array(img) #추가

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white", mask=img_array)
wordcloud = wordcloud.generate(a)

plt.figure() #추가
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
- RAM/Disk Status:** RAM 100%, Disk 100%
- User Profile:** 허나

2.4 워드 클라우드 만들기(명사 빈도수 추출)



2.5 워드 클라우드 만들기 (크롤링)

2.5 워드 클라우드 만들기(크롤링)

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** CO 14_1_5_wordcloud_crawling.ipynb
- File Menu:** 파일 수정 보기 삽입 런타임 도구 도움말 11월 18일에 마지막으로 수정됨
- Toolbar:** 대글 공유 설정 가능
- File Explorer:** 파일 폴더 구조: {x} sample_data NanumGothicExtraBold.ttf heart.png
- Code Cell:** 코딩 모드에서 실행된 코드는 다음과 같습니다.

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from PIL import Image
import numpy as np

import requests
from bs4 import BeautifulSoup

[ ] =====
#크롤링
=====
titles = []
search_word = 'bts'
url = f'https://search.naver.com/search.naver?where=nexearch&sm=top_hty&fbm=1&ie=utf8&query={search_word}'
#print(url)
req = requests.get(url)
html = req.text
#print(html)

soup = BeautifulSoup(html, 'html.parser')
search_result = soup.select_one('.list_news')
news_links = search_result.select('.bx > .news_wrap > .news_area > a')
#print(news_links)

for i in news_links:
    titles.append(i.get_text())
#print(titles)
```

2.5 워드 클라우드 만들기(크롤링)

14_1_5_wordcloud_crawling.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 11월 18일에 마지막으로 수정됨

댓글 공유 설정 가능

파일

코드 텍스트

```
[ ] =====
#워드클라우드
=====
a = '' #리스트의 내용 한 줄로 추출

for x in titles:
    a = a + x
print(a)

img = Image.open('heart.png')
img_array = np.array(img)

wordcloud = WordCloud(font_path="NanumGothicExtraBold.ttf", width=400, height=400, background_color="white", mask=img_array)
wordcloud = wordcloud.generate(a)

plt.figure()
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```

BTS 진 '디 애스트로넷' 누적 102만장 판매…밀리언셀러 등극BTS 진, 첫 솔로 싱글 밀리언셀러 등극입대 앞둔 BTS 진, 바다 배경 화보 공개… “여러 감정 표현” [누구

디스크 85.18 GB 사용 가능

3. QR코드 만들기

QR코드 만들기

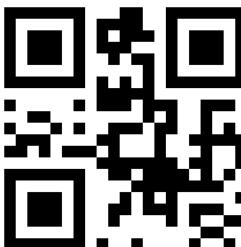
〈Case 1〉

```
!pip install qrcode
!pip install Image

import qrcode

=====
#접속 URL 변경 가능
url = 'google.com'
=====
img = qrcode.make(url)

path = url + '.png'
img.save(path)
```



〈실행 결과〉

import qrcode

=====

#접속 URL 변경 가능
url = 'google.com'

=====

img = qrcode.make(url)

path = url + '.png'
img.save(path)

```
!pip install qrcode
!pip install pillow
```

```
import qrcode
from PIL import Image
=====
```

#원본 이미지

img_bg = Image.open('porori.png')

=====

qr = qrcode.QRCode(box_size=2)

=====

#접속 URL 변경 가능

url = 'https://www.naver.com/'

=====

qr.add_data(url)

qr.make()

img_qr = qr.make_image()

pos = (img_bg.size[0] - img_qr.size[0], img_bg.size[1])

img_bg.paste(img_qr, pos)

=====

#원본 이미지+QR코드

img_bg.save('py.png')

〈Case 2〉



〈원본 이미지〉



〈원본 이미지+QR코드〉



- 파이썬으로 웹 페이지를 크롤링하는 방법을 설명할 수 있다.
- 파이썬으로 워드클라우드를 만들 수 있다.
- 파이썬으로 QR코드를 만들고 URL로 연결할 수 있다.

Q & A

감사합니다