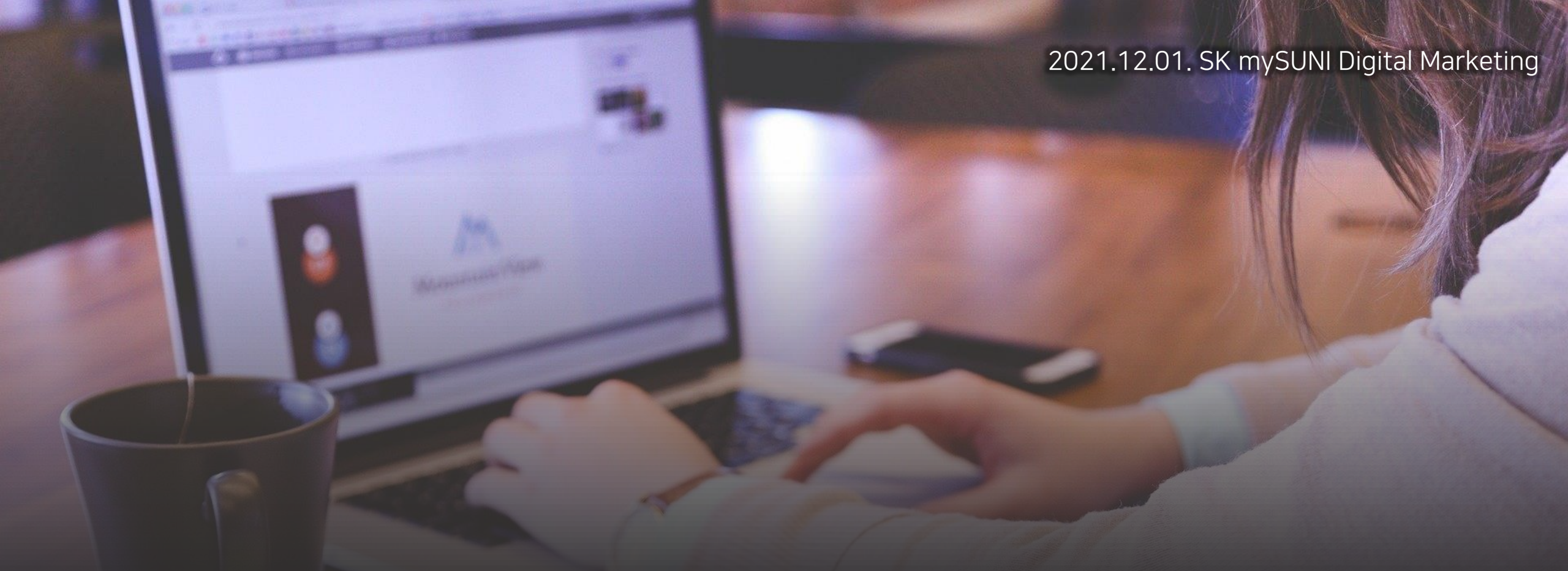


2021.12.01. SK mySUNI Digital Marketing

# 온라인 데이터만으로도 마케터가 될 수 있을까? : BeautifulSoup'만' 활용해서 웹 데이터 수집하기

강태영 (minvv23@.kaist.ac.kr)

**강태영**

minvv23@underscore.kr

## Education

2019.03-2020.08.  M.S. 경영공학 (KAIST)

2014.03.-2019.02  B.A. 사회학/정치학 (서강대학교)

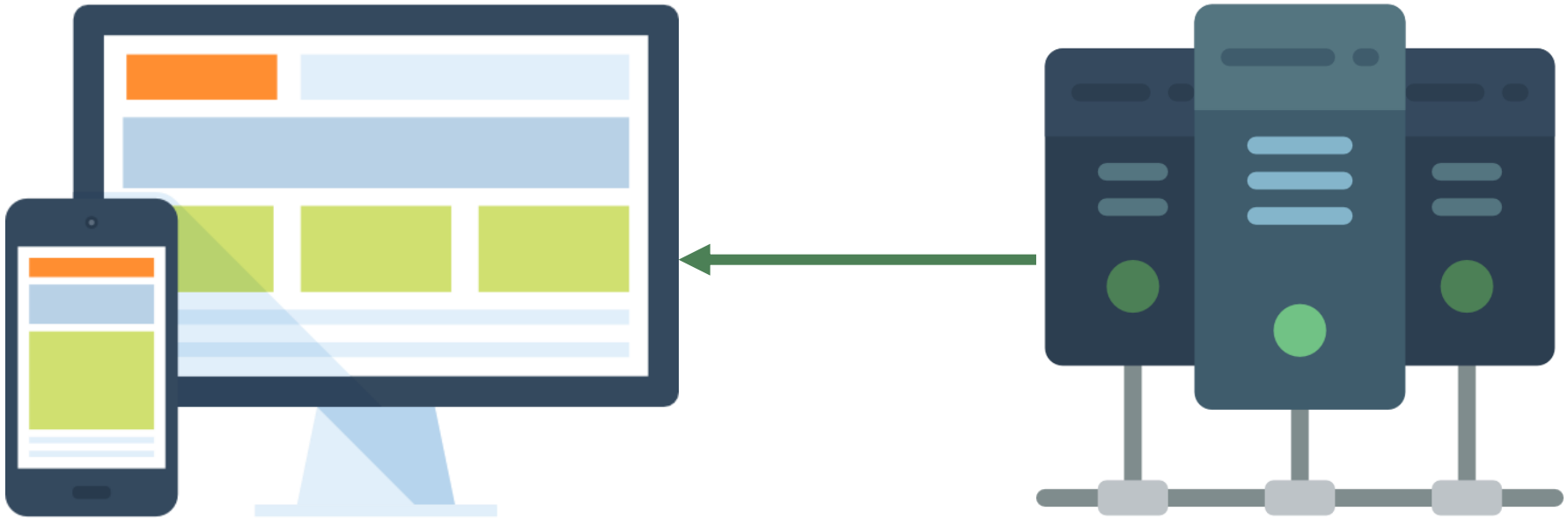
## Job & Activities

- 2018.04.-  데이터 분석 & 미디어 컨텐츠 스타트업 <언더스코어>
- 2020.09.-  KAIST D.LAB Visiting Researcher
- 2021.04.-  MyMusicTaste Data Analyst
- 2020.10.-  KoreaPDS Statistician

## Academic Research & Projects

- AAAI-ICWSM 2021
  "김정은 사망설을 통해 알아보는 온라인 뉴스 댓글에서의 팩트체킹에 대한 당파적 반응 연구"
- SK 신입사원 자기소개서 채점 자동화 모델 개발·평가
- SK Telecom 직원 추천 시스템 및 직무 요약 모델 개발
- BGF리테일 편의점 프랜차이즈 CU 점포 유형화

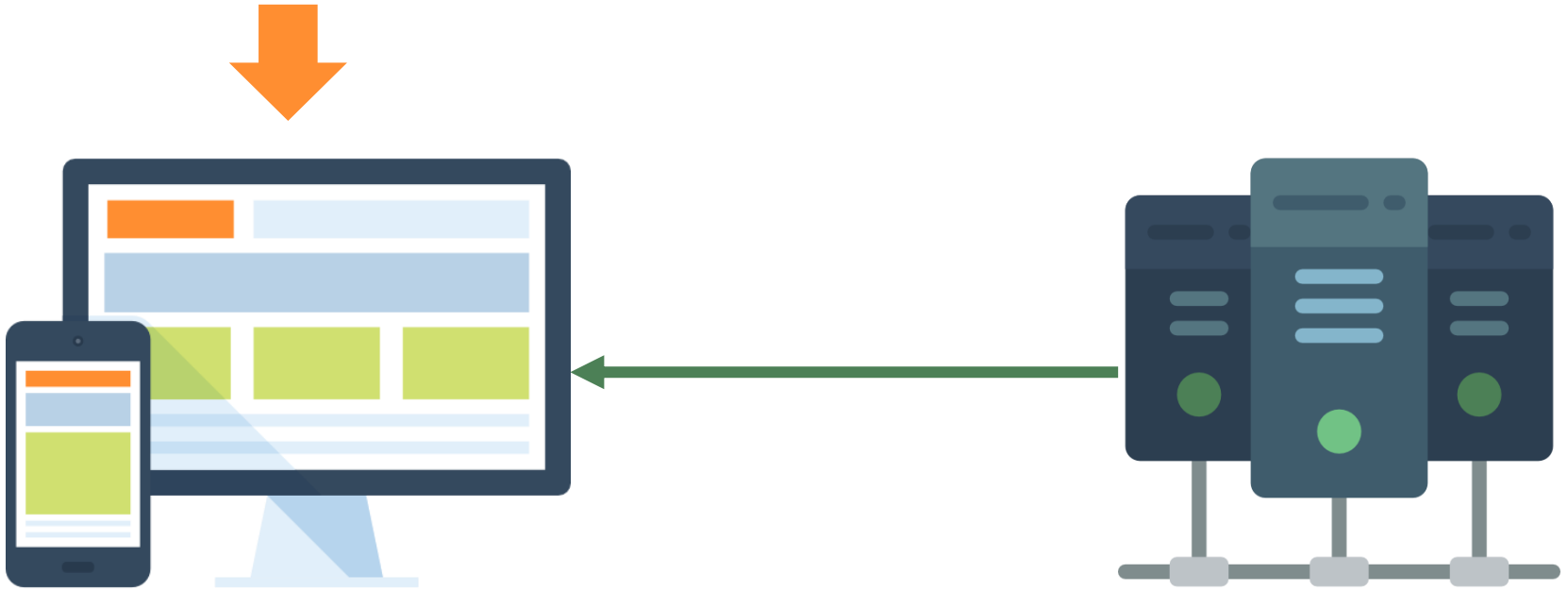
# 웹 크롤링을 하는 방식



**Front-end**

**Back-end**

BeautifulSoup

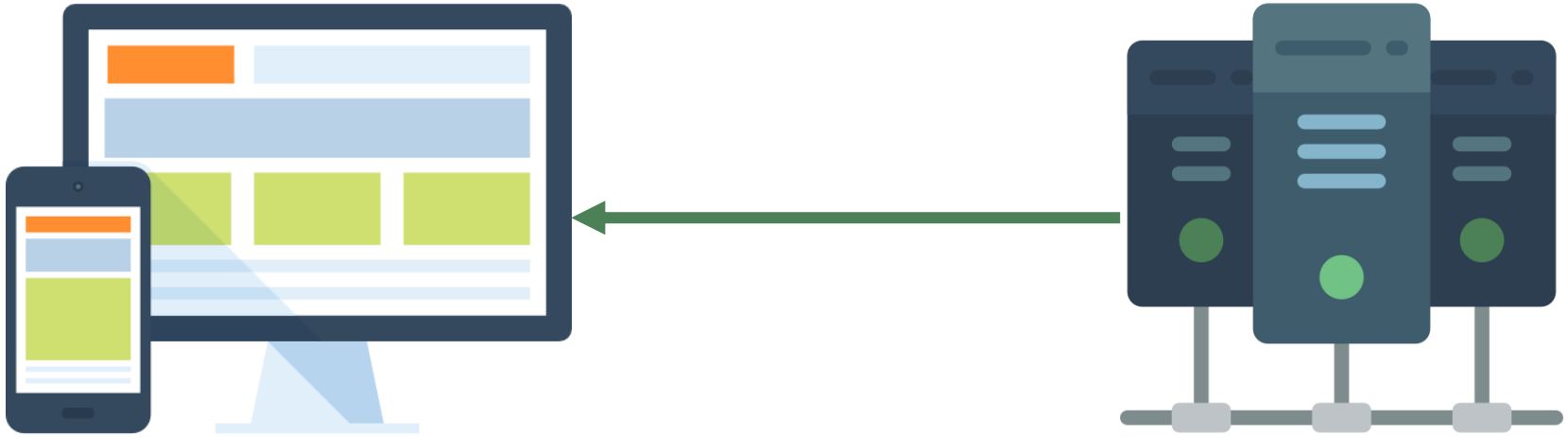


Front-end

Back-end

BeautifulSoup

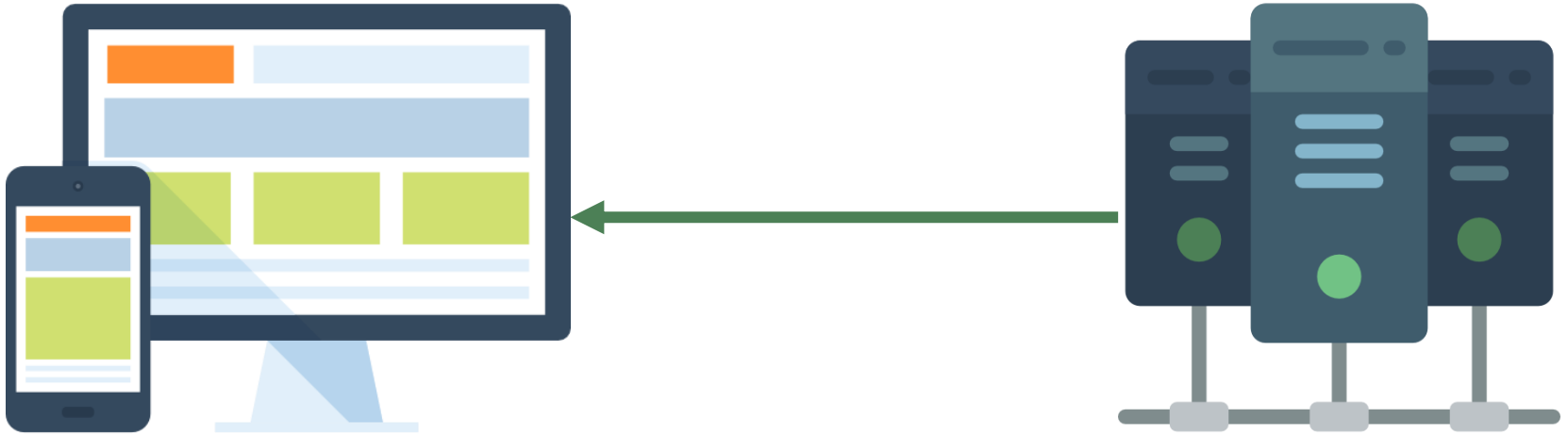
Selenium



Front-end

Back-end

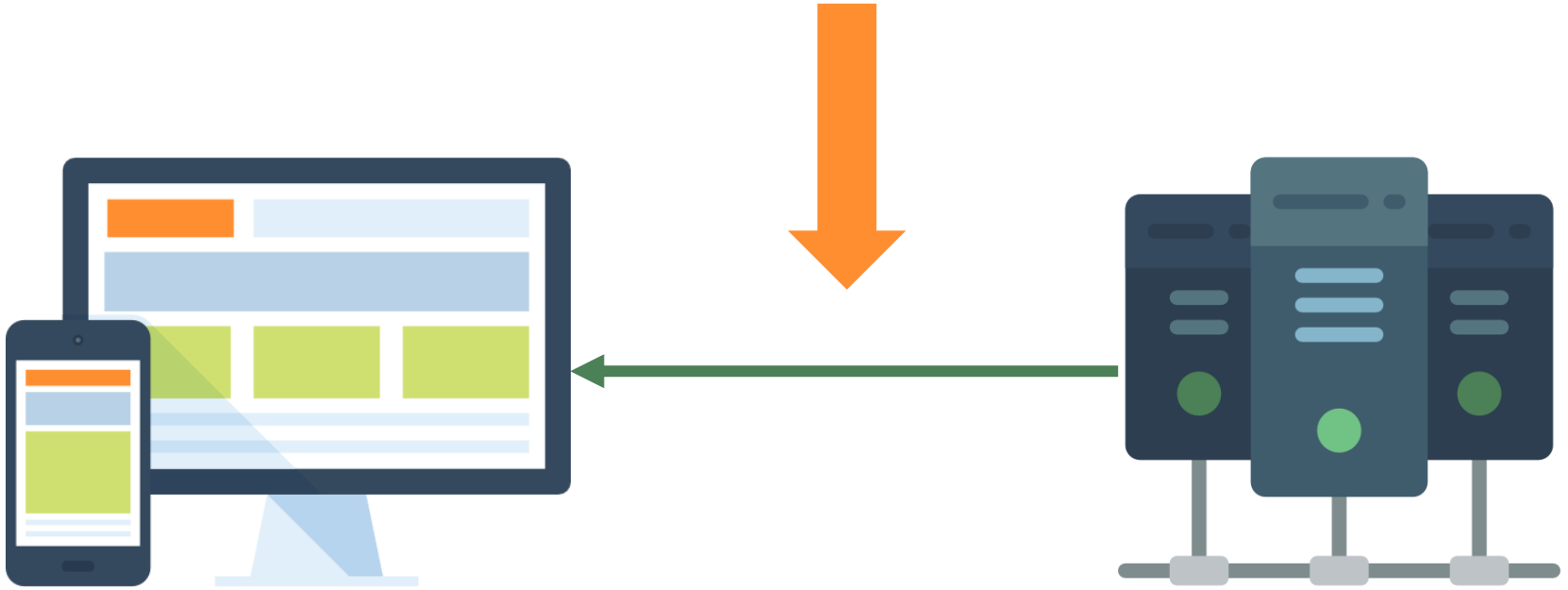
BeautifulSoup



Front-end

Back-end

# BeautifulSoup



Front-end

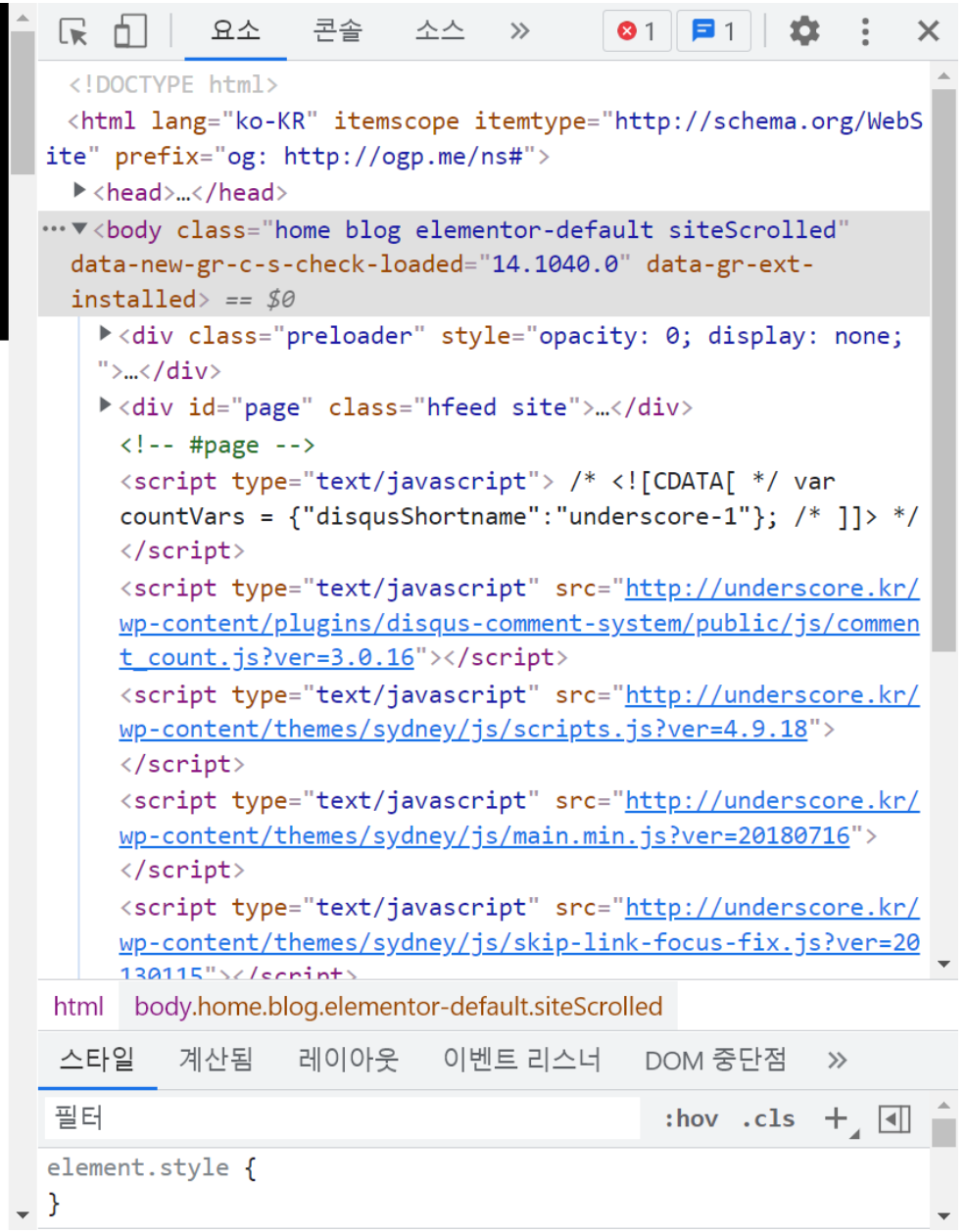
Back-end

KAIST

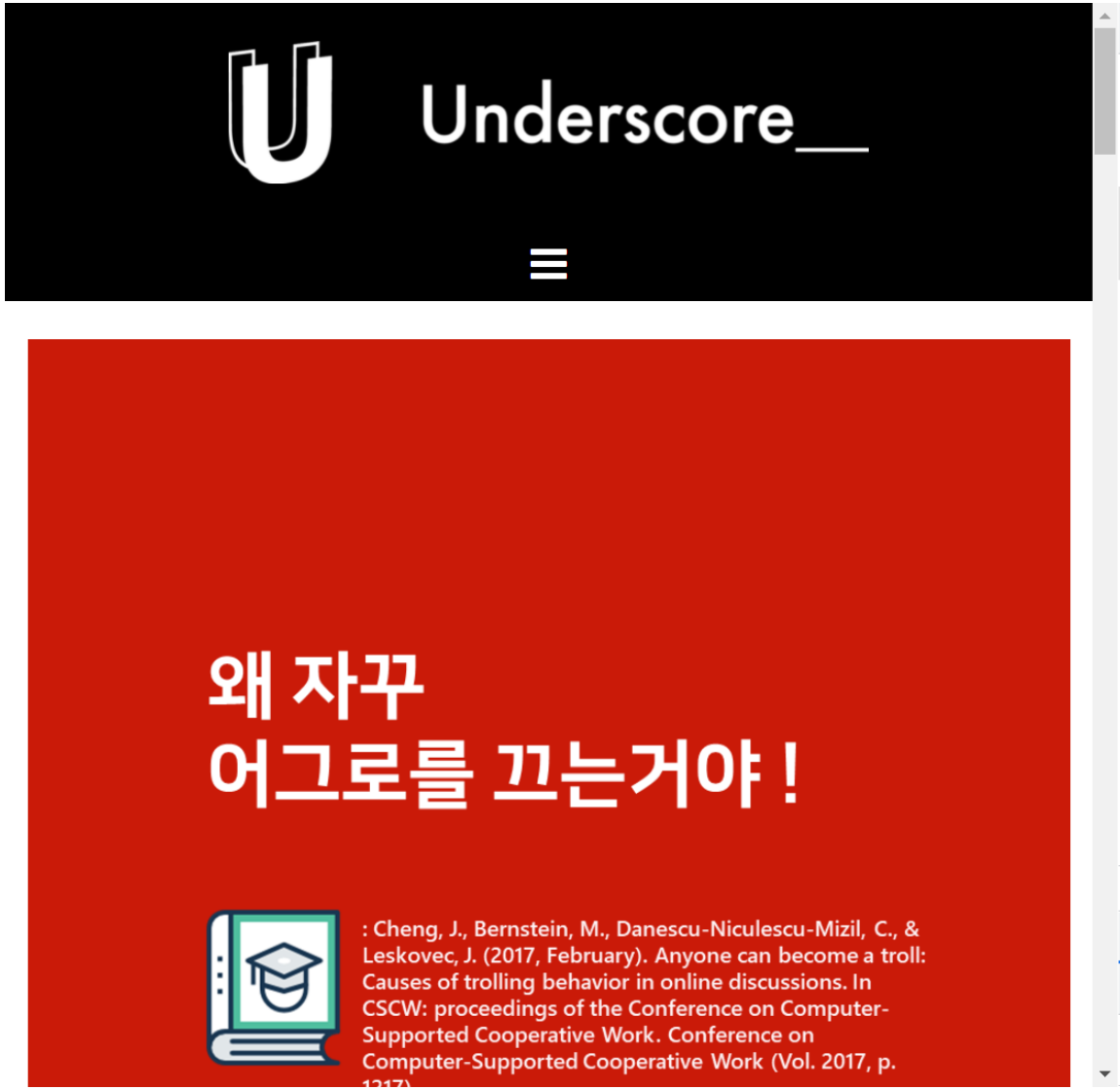
# 웹사이트의 구조

Underscore__

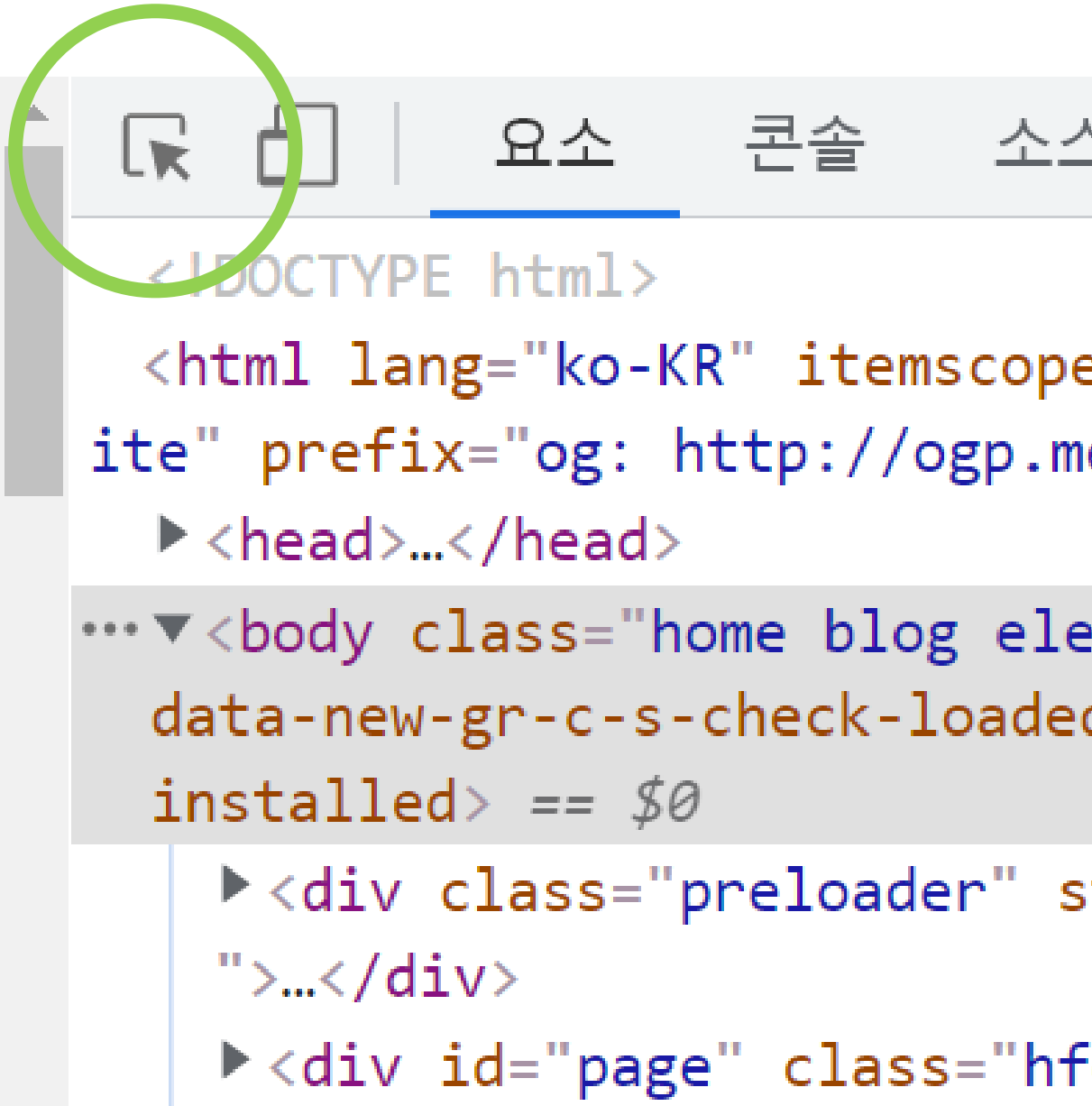
왜 자꾸
어그로를 끄는거야 !
: Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017, February). Anyone can become a troll: Causes of trolling behavior in online discussions. In CSCW: proceedings of the Conference on Computer-Supported Cooperative Work. Conference on Computer-Supported Cooperative Work (Vol. 2017, p.

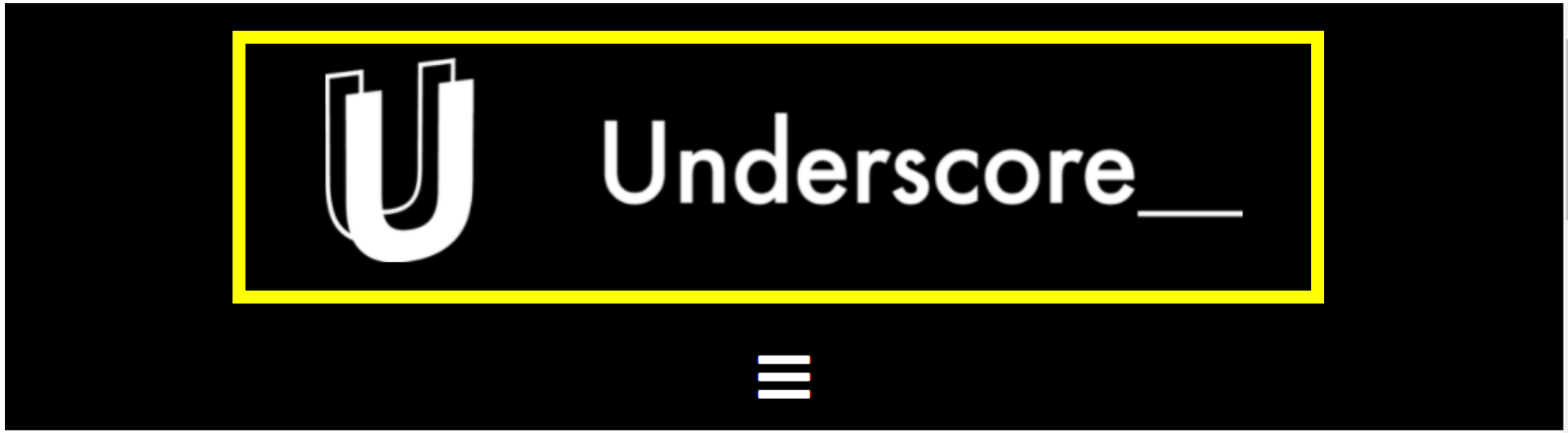

요소
콘솔
소스
<!DOCTYPE html>
<html lang="ko-KR" itemscope itemtype="http://schema.org/WebSite" prefix="og: http://ogp.me/ns#">
<head>…</head>
<body class="home blog elementor-default siteScrolled" data-new-gr-c-s-check-loaded="14.1040.0" data-gr-ext-installed> == $0
…
…
<!-- #page -->
<script type="text/javascript"> /* <![CDATA[ */ var countVars = {"disqusShortname":"underscore-1"}; /* ]]> */
</script>
<script type="text/javascript" src="http://underscore.kr/wp-content/plugins/disqus-comment-system/public/js/comment_count.js?ver=3.0.16"></script>
<script type="text/javascript" src="http://underscore.kr/wp-content/themes/sydney/js/scripts.js?ver=4.9.18">
</script>
<script type="text/javascript" src="http://underscore.kr/wp-content/themes/sydney/js/main.min.js?ver=20180716">
</script>
<script type="text/javascript" src="http://underscore.kr/wp-content/themes/sydney/js/skip-link-focus-fix.js?ver=20
html
body.home.blog.elementor-default.siteScrolled
스타일
계산됨
레이아웃
이벤트 리스너
DOM 중단점
필터
:hov .cls
element.style {
}

# F12 키 누르면 열림

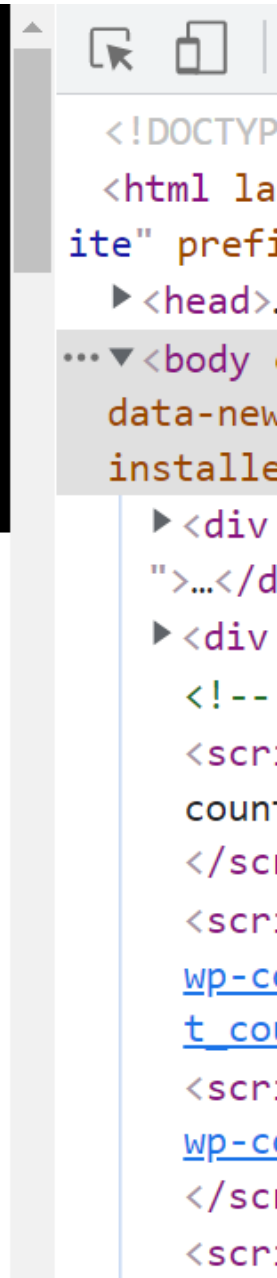

Underscore__
왜 자꾸
어그로를 끄는거야 !
: Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017, February). Anyone can become a troll: Causes of trolling behavior in online discussions. In CSCW: proceedings of the Conference on Computer-Supported Cooperative Work. Conference on Computer-Supported Cooperative Work (Vol. 2017, p.
요소
콘솔
소스
<!DOCTYPE html>
<html lang="ko-KR" itemscope itemtype="http://schema.org/WebSite" prefix="og: http://ogp.me/ns#">
<head>…</head>
<body class="home blog elementor-default siteScrolled" data-new-gr-c-s-check-loaded="14.1040.0" data-gr-ext-installed> == $0
…
…
<!-- #page -->
<script type="text/javascript"> /* <![CDATA[ */ var countVars = {"disqusShortname":"underscore-1"}; /* ]]> */
</script>
<script type="text/javascript" src="http://underscore.kr/wp-content/plugins/disqus-comment-system/public/js/comment_count.js?ver=3.0.16"></script>
<script type="text/javascript" src="http://underscore.kr/wp-content/themes/sydney/js/scripts.js?ver=4.9.18">
</script>
<script type="text/javascript" src="http://underscore.kr/wp-content/themes/sydney/js/main.min.js?ver=20180716">
</script>
<script type="text/javascript" src="http://underscore.kr/wp-content/themes/sydney/js/skip-link-focus-fix.js?ver=20
html
body.home.blog.elementor-default.siteScrolled
스타일
계산됨
레이아웃
이벤트 리스너
DOM 중단점
필터
:hov
.cls
element.style {
}

core___


요소
콘솔
<!DOCTYPE html>
<html lang="ko-KR" itemscope
ite" prefix="og: http://ogp.m
<head>…</head>
<body class="home blog ele
data-new-gr-c-s-check-loaded
installed> == $0
<div class="preloader" s
">…
<div id="page" class="hf


Underscore__

왜 자꾸


<!DOCTYP
<html la
ite" prefi
<head>
<body
data-new
installe
<div
">…</d
<div
<!--
<scr
count
</sc
<scr
wp-c
t_co
<scr
wp-c
</sc
<scr

Underscore__

왜 자꾸

<!DOCTYP
<html la
ite" prefi
<head>
<body
data-new
installe
<div
">…</d
<div
<!--
<scr
count
</sc
<scr
wp-c
t_co
<scr
wp-c
</sc
<scr

Underscore_

왜 자꾸
어그로를 끄는거야 !
: Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017, February). Anyone can become a troll: Causes of trolling behavior in online discussions. In CSCW: proceedings of the Conference on Computer-Supported Cooperative Work. Conference on Computer-Supported Cooperative Work (Vol. 2017, p.

요소
콘솔
소스
<!DOCTYPE html>
<html lang="ko-KR" itemscope itemtype="http://schema.org/WebSite" prefix="og: http://ogp.me/ns#">
<head>…</head>
<body class="home blog elementor-default siteScrolled" data-new-gr-c-s-check-loaded="14.1040.0" data-gr-ext-installed> == $0
…
…
<script type="text/javascript"> /* <![CDATA[ */ var countVars = {"disqusShortname":"underscore-1"}; /* ]]> */ </script>
<script type="text/javascript" src="http://underscore.kr/wp-content/plugins/disqus-comment-system/public/js/comment_count.js?ver=3.0.16"></script>
<script type="text/javascript" src="http://underscore.kr/wp-content/themes/sydney/js/scripts.js?ver=4.9.18"></script>
<script type="text/javascript" src="http://underscore.kr/wp-content/themes/sydney/js/main.min.js?ver=20180716"></script>
<script type="text/javascript" src="http://underscore.kr/wp-content/themes/sydney/js/skip-link-focus-fix.js?ver=20
html
body.home.blog.elementor-default.siteScrolled
스타일
계산됨
레이아웃
이벤트 리스너
DOM 중단점
필터
:hov .cls
element.style {
}

# 태그

```
▶ <div id="page" class="hfeed site">…</div>
  <!-- #page -->
```

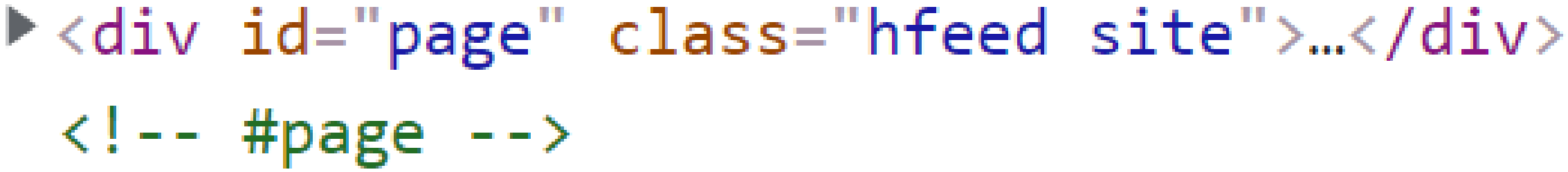
## 속성

```
▶ <div id="page" class="hfeed site">…</div>
  <!-- #page -->
```

## 속성값

```
▶ <div id="page" class="hfeed site">…</div>
  <!-- #page -->
```

# 속성



속성값

```
▶ <div id="page" class="hfeed site">…</div>
  <!-- #page -->
```

KAIST

# 실습하러 갑시다

# 감사합니다.

강태영 (minvv23@.underscore.kr)