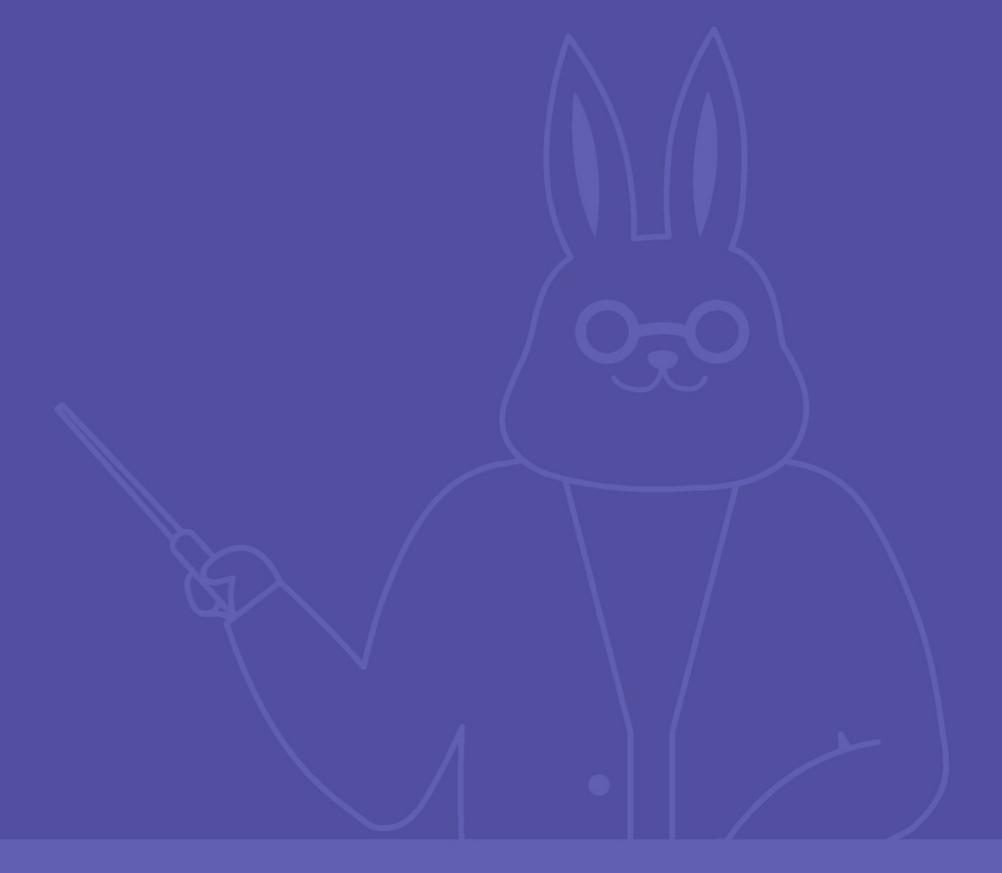


파에썬 크롤링입문

04 웹 스크래핑 프로젝트

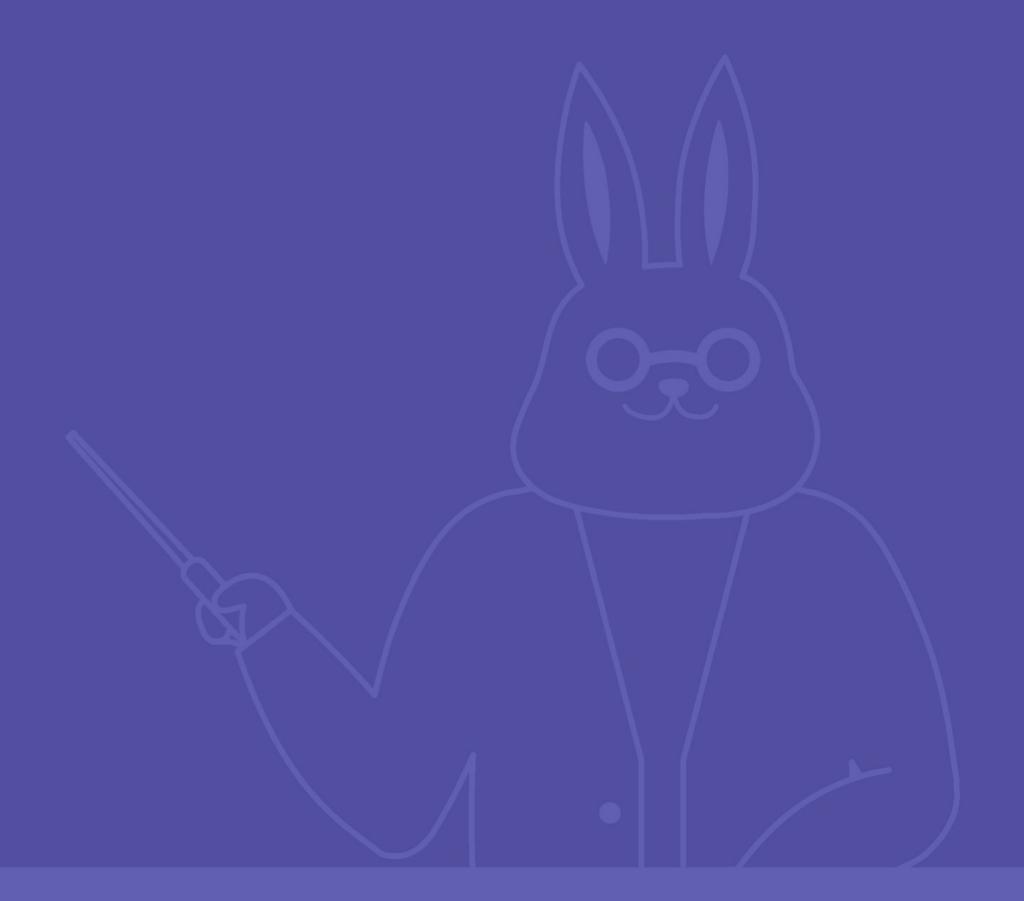




- 01. 프로젝트에 관하여
- 02. 첫 번째 프로젝트
- 03. 두 번째 프로젝트
- 04. 세 번째 프로젝트

01

프로젝트에 관하여





지금까지 배운 내용을 토대로 진행하는 실습 위주의 수업

01 프로젝트에 관하여

❷ 어떤 프로젝트를 진행하나요?



ıs: Fo	rn	ns,	Se	ear	chi	n	g ar	nd	Pag	gina	atio	n 2	5 item
ase of N k at how paginate	/ pag	ginat	ion a	nd s	earch					_			
show you h	ow to	scrap	e this	page.									Data v
			Se	earch	for Tear	ms:	Sear	ch foi	r Teams		Search	1	
	J		770.7						= -				
Winnipe	eg Je	ts					1990		26		43		
Boston	Bruii	ns					1991		36	;	32		
Buffalo	Sabr	es					1991		31		37		
Calgary			1991		31		37						
Chicago	o Bla	ckha	wks				1991		36	;	29		
1	2	3	4	5	6	7	8	9	10	11	12	13	14

This Site FAQ Sandbox **Lessons** Winning Films: AJAX and Javascript 87 items a bunch of great films. Learn how content is added to the page asynchronously with bk for ways that you can tell visually when a site is loading content with AJAX. Then, by AJAX requests and scrape them. Data via https://en.wikipedia.or eo lessons that show you how to scrape this page. **Choose a Year to View Films** 2014 2013 2012 2011 2010 <u>2015</u> Lessons and Videos © Hartley Brody 2018

간단한 페이지

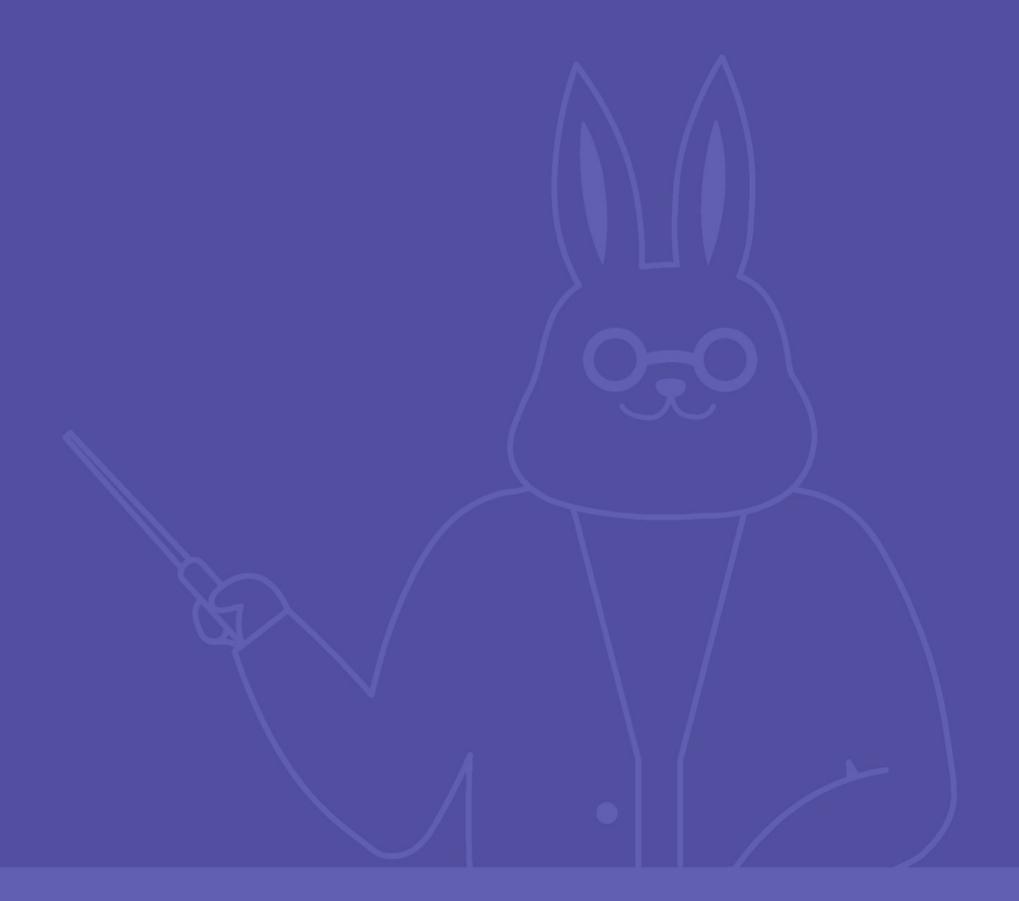
검색 & Pagination

JS를 통한 동적 렌더링

☑ 프로젝트 수업의 목표

나만의 Scraper를 만들 수 있는 실력 함양

첫번째프로젝트



02 첫 번째 프로젝트

❷ 웹 페이지 소개

Area (km²): 468.0

Antigua and Barbuda

Capital: St. John's Population: 86754 Area (km²): 443.0



Capital: Yerevan
Population: 2968000
Area (km²): 29800.0

Argentina

Capital: Buenos Aires Population: 41343201 Area (km²): 2766890.0

Australia

Capital: Canberra Population: 21515754 Area (km²): 7686850.0

Azerbaijan

Capital: Baku
Population: 8303512
Area (km²): 86600.0

Panaladach

Area (km²): 82880.0

Anguilla

Capital: The Valley Population: 13254 Area (km²): 102.0

Angola

Capital: Luanda
Population: 13068161
Area (km²): 1246700.0

American Samoa

Capital: Pago Pago Population: 57881 Area (km²): 199.0

Aruba

Capital: Oranjestad Population: 71566 Area (km²): 193.0

Sosnia and Herzegovina

Capital: Sarajevo Population: 4590000 Area (km²): 51129.0

Dolaium

Area (km²): 647500.0

***** Albania

Capital: Tirana
Population: 2986952
Area (km²): 28748.0

Antarctica

Capital: None
Population: 0
Area (km²): 1.4E7

Austria

Capital: Vienna
Population: 8205000
Area (km²): 83858.0

Åland

Capital: Mariehamn
Population: 26711
Area (km²): 1580.0

♥ Barbados

Capital: Bridgetown Population: 285653 Area (km²): 431.0

Durking Esca

02 첫 번째 프로젝트

❷ 프로젝트 개요

Andorra

Capital: Andorra la Vella Population: 84000 Area (km²): 468.0

Antigua and Barbuda

Capital: St. John's Population: 86754 Area (km²): 443.0

Armenia

Capital: Yerevan Population: 2968000 Area (km²): 29800.0

Argentina

Capital: Buenos Aires Population: 41343201 Area (km²): 2766890.0

** Australia

Capital: Canberra Population: 21515754 Area (km²): 7686850.0

Azerbaijan

Capital: Baku Population: 8303512 Area (km²): 86600.0

Bangladesh

Capital: Dhaka
Population: 156118464
Area (km²): 144000.0

United Arab Emirates

Capital: Abu Dhabi Population: 4975593 Area (km²): 82880.0

Anguilla Anguilla

Capital: The Valley Population: 13254
Area (km²): 102.0

Angola

Capital: Luanda
Population: 13068161
Area (km²): 1246700.0

American Samoa

Capital: Pago Pago Population: 57881 Area (km²): 199.0

Aruba

Capital: Oranjestad Population: 71566 Area (km²): 193.0

Bosnia and Herzegovina

Capital: Sarajevo Population: 4590000 Area (km²): 51129.0

Belgium

Capital: Brussels
Population: 10403000
Area (km²): 30510.0

Afghanistan

Capital: Kabul Population: 29121286 Area (km²): 647500.0

Albania

Capital: Tirana Population: 2986952 Area (km²): 28748.0

Marctica (Marctica)

Capital: None Population: 0 Area (km²): 1.4E7

Austria

Capital: Vienna Population: 8205000 Area (km²): 83858.0

Åland

Capital: Mariehamn Population: 26711 Area (km²): 1580.0

■ Barbados

Capital: Bridgetown Population: 285653 Area (km²): 431.0

Burkina Faso

Capital: Ouagadougou Population: 16241811 Area (km²): 274200.0

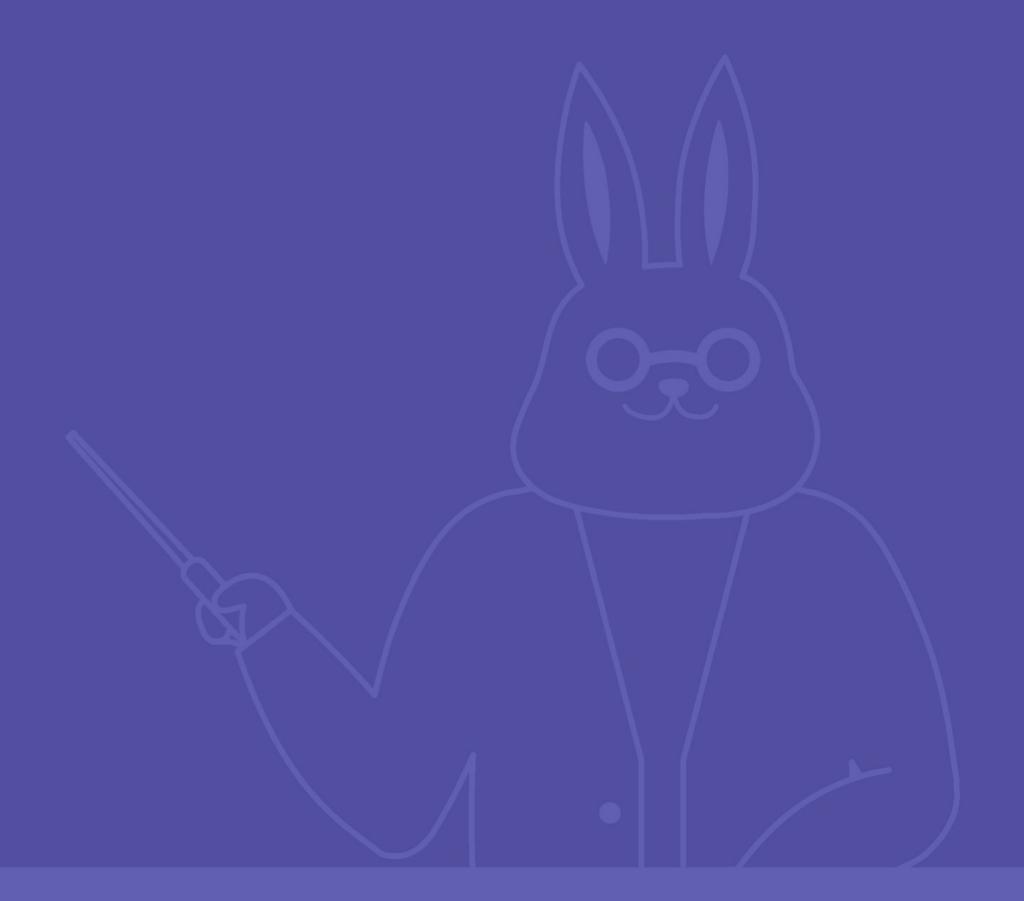
웹 페이지 특징

- 브라우저 진입 시, 모든 데이터가 한 번에 로딩 된다.
- 대량의 데이터가 존재

목표

- 대량의 데이터가 로딩되기에, Wait 기능 필수
- 데이터가 많을 때, 정제하는 방법 학습

두 번째 프로젝트



03 두 번째 프로젝트

❷ 웹 페이지 소개

			Search for To	eams: Search for	Teams	Search		
Team Name	Year	Wins	Losses	OT Losses	Win %	Goals For (GF)	Goals Against (GA)	+/-
Boston Bruins	1990	44	24		0.55	299	264	35
Buffalo Sabres	1990	31	30		0.388	292	278	14
Calgary Flames	1990	46	26		0.575	344	263	81
Chicago Blackhawks	1990	49	23		0.613	284	211	73
Detroit Red Wings	1990	34	38		0.425	273	298	-25
Edmonton Oilers	1990	37	37		0.463	272	272	0
Hartford Whalers	1990	31	38		0.388	238	276	-38
Los Angeles Kings	1990	46	24		0.575	340	254	86
Minnesota North Stars	1990	27	39		0.338	256	266	-10
Montreal Canadiens	1990	39	30		0.487	273	249	24
New Jersey Devils	1990	32	33		0.4	272	264	8
New York Islanders	1990	25	45		0.312	223	290	-67
New York Rangers	1990	36	31		0.45	297	265	32

1990년 ~ 2011년 동안 미국 하키 팀의 성적

❷ 프로젝트 개요

			Search for Teams:	Search for Teams	Search	
	Year	Wins	Losses OT L	osses Win %	Goals For (GF)	Goals Against (GA)
	1990	44	24	0.55	299	264
	1990	31	30	0.388	292	278
	1990	46	26	0.575	344	263
vks	1990	49	23	0.613	284	211
S	1990	34	38	0.425	273	298
	1990	37	37	0.463	272	272
	1990	31	38	0.388	238	276
S	1990	46	24	0.575	340	254
Stars	1990	27	39	0.338	256	266
ns	1990	39	30	0.487	273	249
3	1990	32	33	0.4	272	264
rs	1990	25	45	0.312	223	290
3	1990	36	31	0.45	297	265
S	1990	33	37	0.412	252	267
ns	1990	41	33	0.512	342	305

웹 페이지 특징

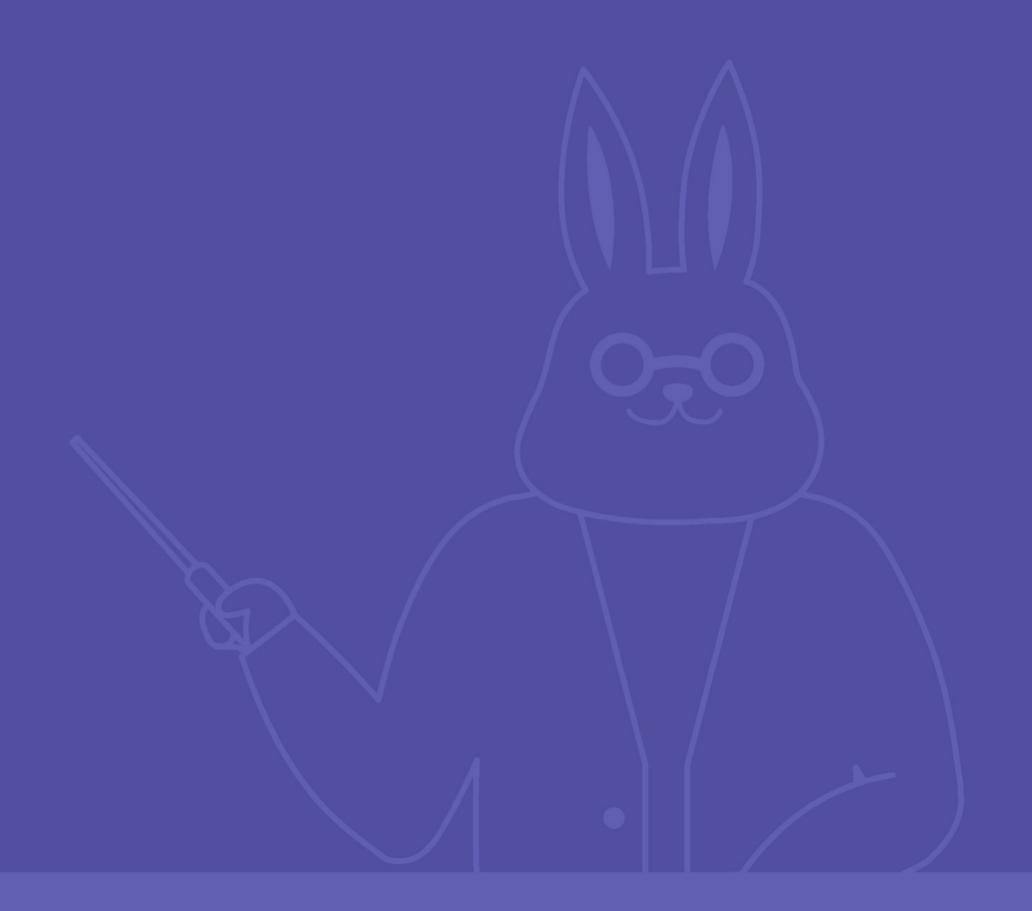
- 검색 기능이 있음
- Pagination을 통해 모든 데이터가 한 페이지에 담기지 않음

목표

- 검색 기능을 활용한 스크래핑을 할 수 있음
- 여러 페이지에 걸쳐 정보 추출

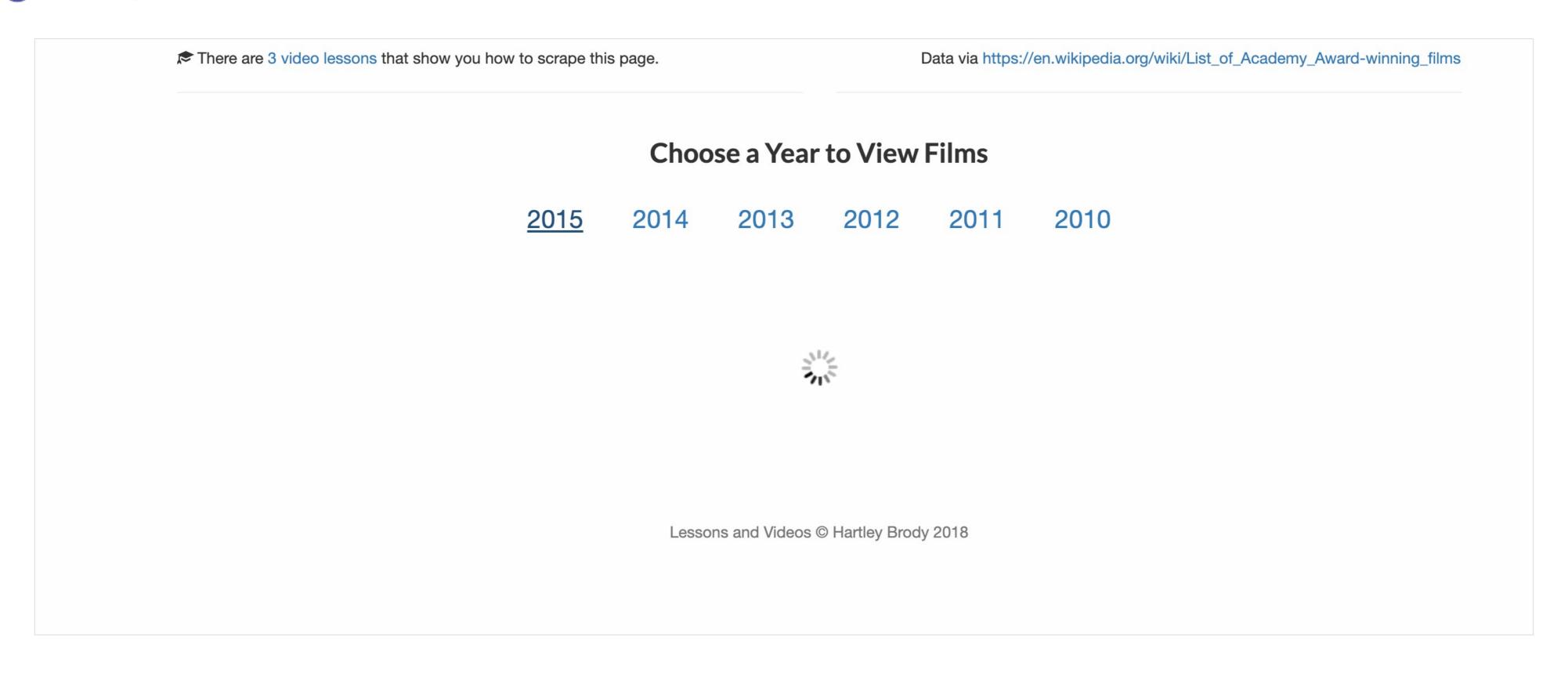
04

세번째프로젝트



04 세 번째 프로젝트

❷ 웹 페이지 소개



연도별, 작품별 오스카 상 수상 목록

❷ 프로젝트 개요

bunch of great films. Learn how content is added to the page asynchronously with Javascript and how for ways that you can tell visually when a site is loading content with AJAX. Then, browse through you AX requests and scrape them.

essons that show you how to scrape this page.

Data via https://en.wikipedia.org/wiki/List_of_Academy_A

Choose a Year to View Films

<u>2015</u> 2014 2013 2012 2011 2010



Lessons and Videos © Hartley Brody 2018

웹 페이지 특징

- 처음엔 데이터가 전혀 없음
- 버튼을 누르면 데이터를 ajax 통신으로 가져
 오고, 이를 웹페이지에 보여줌

목표

 버튼을 눌러야만 표시되는 데이터를 추출할 수 있음

크레딧

/* elice */

코스 매니저 임승연

콘텐츠 제작자 신용기

강사 신용기

감수자 장석준

디자이너 강혜정

연락처

TEL

070-4633-2015

WEB

https://elice.io

E-MAIL

contact@elice.io

