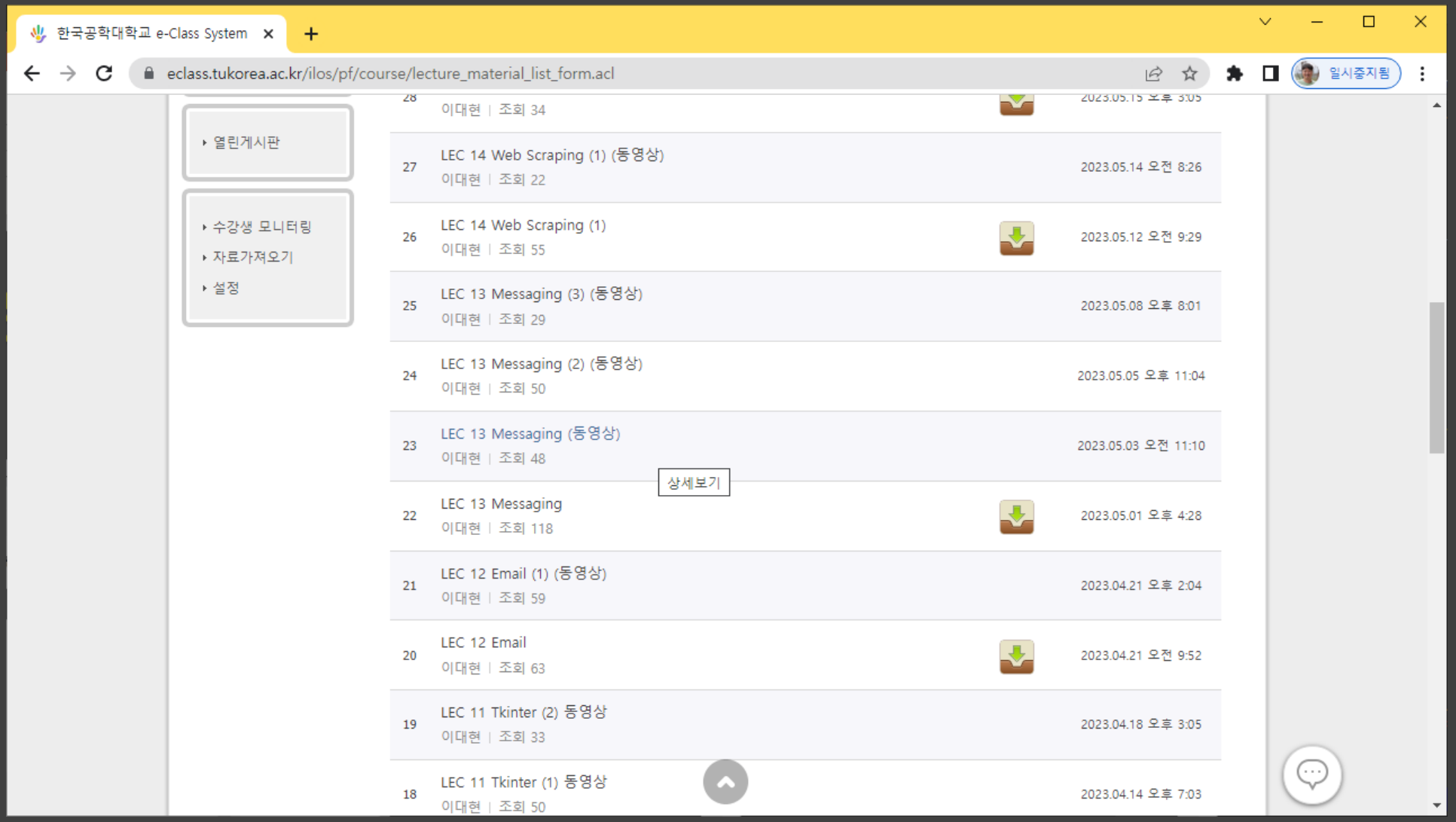
제 14 강 Web Scraping (3)

© 이대현

과업 #4. eclass 에서 강의 자료 일괄 다운로드



© 이대현

로그인

```
from playwright.sync_api import sync_playwright
import time

url = 'https://eclass.tukorea.ac.kr/ilos/main/member/login_form.acl'
user_agent = 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/125.0.0.0 Safari/537.36'

p = sync_playwright().start()
# with sync_playwright() as p:
browser = p.chromium.launch(headless=False).new_context(
    user_agent=user_agent,
    viewport={'width':1024, 'height': 1024}
)
page = browser.new_page()
page.goto(url)
page.locator('#usr_id').fill('dustinlee')
page.locator('#usr_pwd').fill('XXXXXXXX')
page.locator('#login_btn').click()
time.sleep(20)
browser.close()
p.stop()
```

© 이대현

강의 자료 페이지 이동

```
page.locator('em[title="스크립트언어"]').click()
page.click('#menu_lecture_material') # locator 필요없이 바로 클릭
page.reload()
```

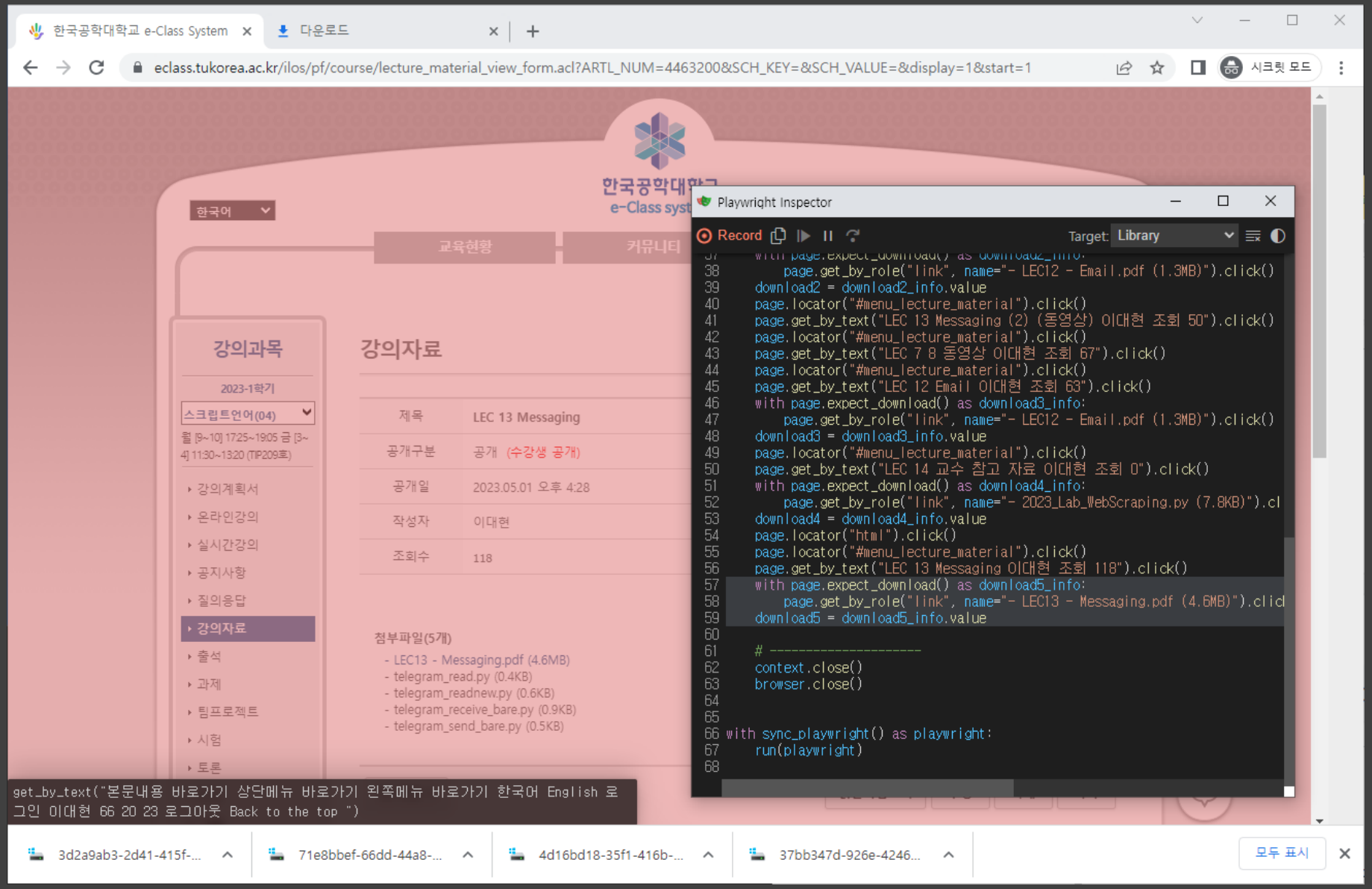
© 이대현

개별 게시물 이동

```
elms = page.locator('.site-link').all()
for e in elms[:3]:
    e.click()
    time.sleep(1)
    page.click('#myform > div.bbs-rbutton04 > div:nth-child(4)')
```

© 이대현

자동화 추출 : playwright codegen



© 이대현

전체 접속 및 다운로드 과정 녹화

Playwright Inspector

Record | Target: Pytest

```python
import re
from playwright.sync_api import Page, expect


def test_example(page: Page) -> None:
    page.goto("https://eclass.tukorea.ac.kr/ilos/main/member/login_form.acl")
    page.get_by_role("textbox", name="ID").click()
    page.get_by_role("textbox", name="ID").fill("dustinlee")
    page.get_by_role("textbox", name="ID").press("Tab")
    page.get_by_role("textbox", name="Password").fill("[redacted]")
    page.get_by_role("textbox", name="Password").press("Enter")
    page.get_by_title("Script Languages", exact=True).click()
    page.locator("#menu_lecture_material").click()
    page.get_by_text("LEC 14 Web Scraping (2) (동영상 링크 추가) 이대현 Search").click()
    with page.expect_download() as download_info:
        page.get_by_role("link", name="- LEC14 - Web Scraping (2).").click()
    download = download_info.value

```

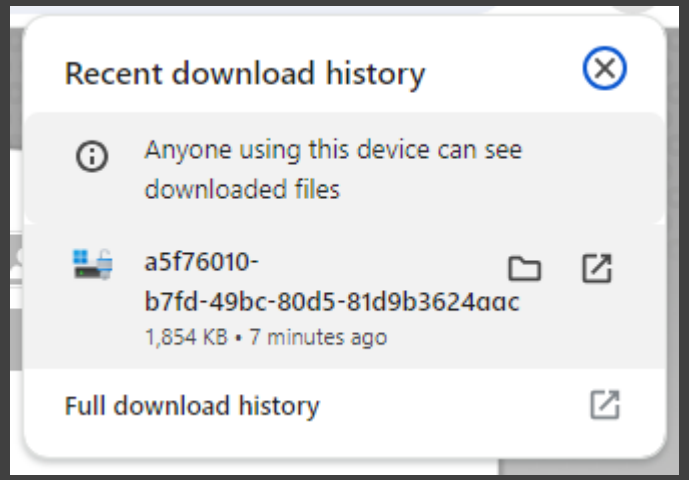
Locator | Log

```
locator("#header").get_by_role("link", name="Log In")
```

© 이대현

파일 다운로드 분석

```
with page.expect_download() as download_info:
    page.get_by_role("link", name="- LEC14 - Web Scraping (2).").click()
download = download_info.value
download.save_as(download.suggested_filename)
print(download)
```



© 이대현

파일 다운로드

```
elms = page.locator('a.site-link').all()
for e in elms[:3]: # 상위 3개만
    e.click()

    for l in page.locator('a.site-link').all():
        with page.expect_download() as download_info:
            l.click()
        download = download_info.value
        download.save_as(download.suggested_filename)

    time.sleep(1)
    page.click('#myform > div.bbs-rbutton04 > div:nth-child(4)')
```

© 이대현

네이버 로그인

```
from playwright.sync_api import sync_playwright
import time


url = 'https://nid.naver.com/nidlogin.login?mode=form&url=https://www.naver.com/'
p = sync_playwright().start()

browser = p.chromium.launch(headless=False).new_context(
    viewport={'width': 1920, 'height': 1024}
)

page = browser.new_page()
page.goto(url)
print(page.title())


page.get_by_role('textbox', name='아이디').fill('*******')
page.get_by_role('textbox', name='비밀번호').fill('*******')
page.get_by_role('button', name='로그인').click()
page.get_by_text('등록', exact=True).click()
```

© 이대현