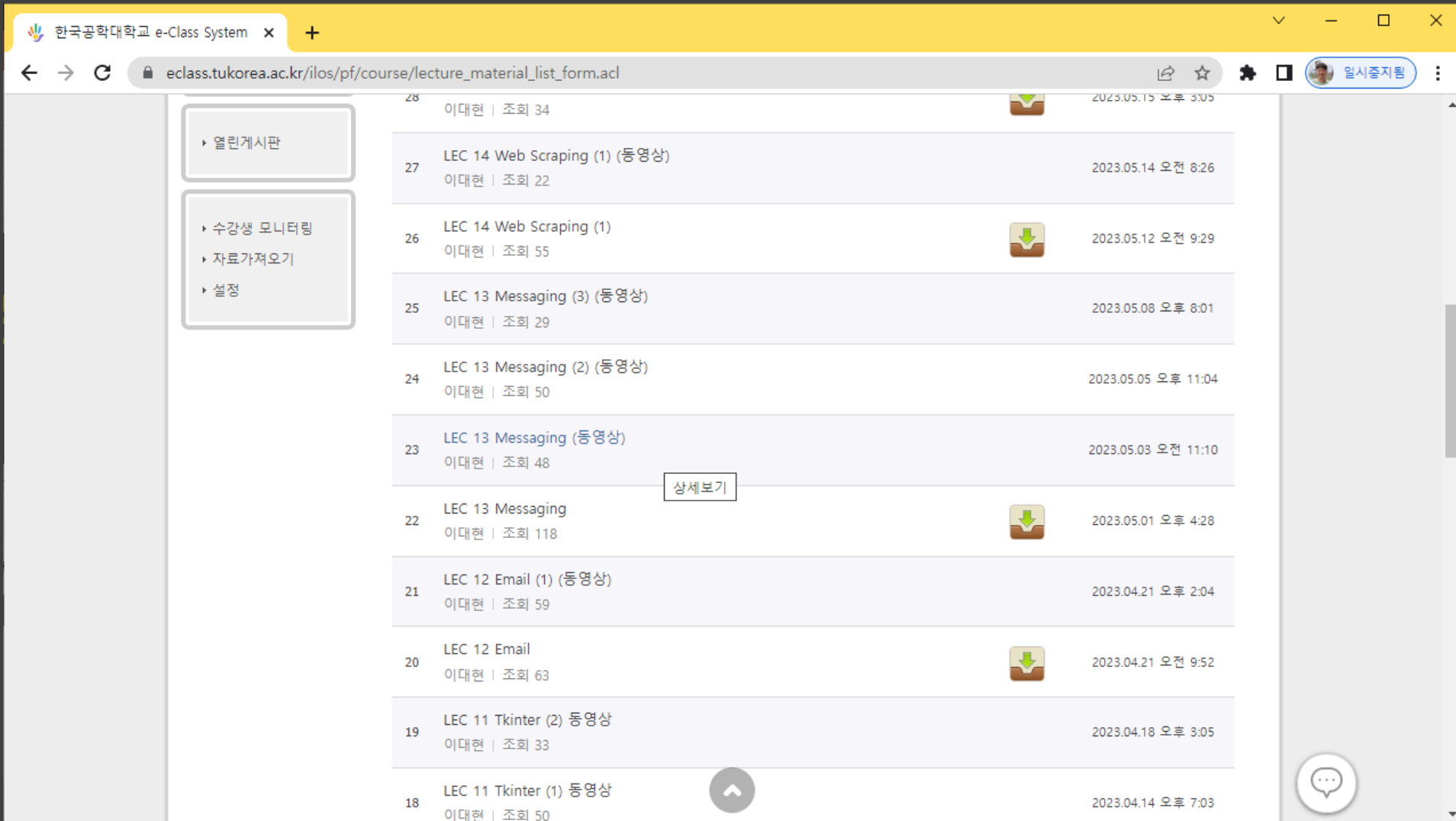


# 제 14 강 Web Scraping (3)

# 과업 #4. eclass 에서 강의 자료 일괄 다운로드



The screenshot shows the 'e-class' system interface for Hanyang University. The browser address bar displays 'eclass.tukorea.ac.kr/ilos/pf/course/lecture\_material\_list\_form.acl'. The left sidebar contains navigation links: '열린게시판', '수강생 모니터링', '자료가져오기', and '설정'. The main content area lists lecture materials with columns for item number, title, lecturer/view count, download status, and date. A '상세보기' button is visible next to item 23.

번호	제목	강사	조회	다운로드	작성일
28	이대현   조회 34				2023.05.15 오후 3:05
27	LEC 14 Web Scraping (1) (동영상)	이대현	조회 22		2023.05.14 오전 8:26
26	LEC 14 Web Scraping (1)	이대현	조회 55		2023.05.12 오전 9:29
25	LEC 13 Messaging (3) (동영상)	이대현	조회 29		2023.05.08 오후 8:01
24	LEC 13 Messaging (2) (동영상)	이대현	조회 50		2023.05.05 오후 11:04
23	LEC 13 Messaging (동영상)	이대현	조회 48		2023.05.03 오전 11:10
22	LEC 13 Messaging	이대현	조회 118		2023.05.01 오후 4:28
21	LEC 12 Email (1) (동영상)	이대현	조회 59		2023.04.21 오후 2:04
20	LEC 12 Email	이대현	조회 63		2023.04.21 오전 9:52
19	LEC 11 Tkinter (2) 동영상	이대현	조회 33		2023.04.18 오후 3:05
18	LEC 11 Tkinter (1) 동영상	이대현	조회 50		2023.04.14 오후 7:03

# 로그인

```
from playwright.sync_api import sync_playwright
import time

url = 'https://eclass.tukorea.ac.kr/ilos/main/member/login_form.ac1'
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()
```

# 강의 자료 페이지 이동

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

# 개별 게시물 이동

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

# 자동화 추출 : playwright codegen

The screenshot shows a web browser window displaying the e-Class System interface. The URL is `eclass.tukorea.ac.kr/ilos/pf/course/lecture_material_view_form.acl?ARTL_NUM=4463200&SCH_KEY=&SCH_VALUE=&display=1&start=1`. The page title is "한국공학대학교 e-Class System". The interface includes a sidebar with "강의과목" (Lecture Course) and "강의자료" (Lecture Material) sections. The "강의자료" section shows a table of materials for "LEC 13 Messaging".

제목	LEC 13 Messaging
공개구분	공개 (수강생 공개)
공개일	2023.05.01 오후 4:28
작성자	이대현
조회수	118

Below the table, there is a section for "첨부파일(5개)" (Attachments (5 files)):

- LEC13 - Messaging.pdf (4.6MB)
- telegram\_read.py (0.4KB)
- telegram\_readnew.py (0.6KB)
- telegram\_receive\_bare.py (0.9KB)
- telegram\_send\_bare.py (0.5KB)

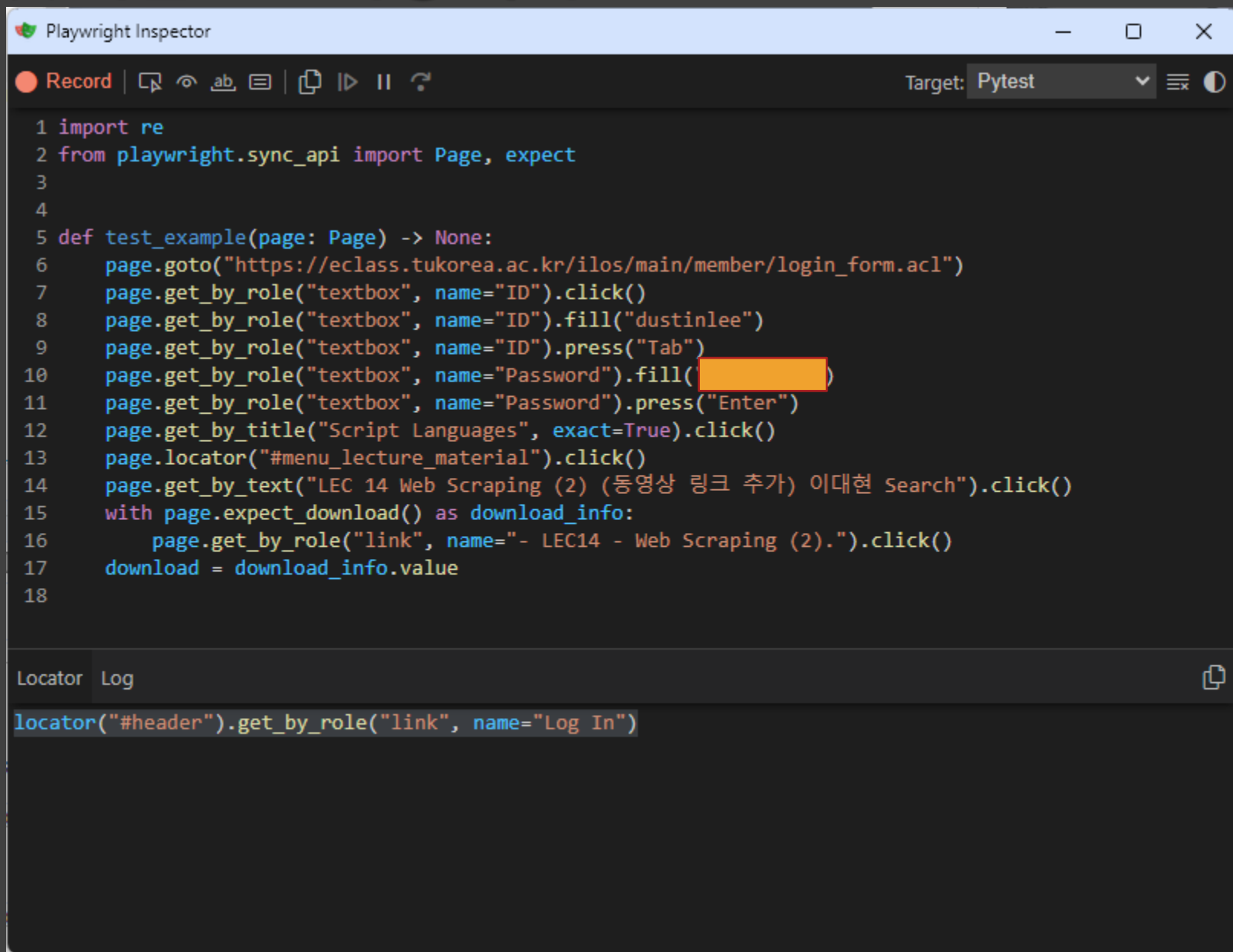
On the right side, the Playwright Inspector window is open, showing the recorded code for the current session. The code includes actions like clicking on links and downloading files.

```
Record
Target: Library
37 with page.expect_download() as download2_info:
38     page.get_by_role("link", name="LEC12 - Email.pdf (1.3MB)").click()
39     download2 = download2_info.value
40     page.locator("#menu_lecture_material").click()
41     page.get_by_text("LEC 13 Messaging (2) (동영상) 이대현 조회 50").click()
42     page.locator("#menu_lecture_material").click()
43     page.get_by_text("LEC 7 8 동영상 이대현 조회 67").click()
44     page.locator("#menu_lecture_material").click()
45     page.get_by_text("LEC 12 Email 이대현 조회 63").click()
46     with page.expect_download() as download3_info:
47         page.get_by_role("link", name="LEC12 - Email.pdf (1.3MB)").click()
48         download3 = download3_info.value
49         page.locator("#menu_lecture_material").click()
50         page.get_by_text("LEC 14 교수 참고 자료 이대현 조회 0").click()
51         with page.expect_download() as download4_info:
52             page.get_by_role("link", name="2023_Lab_WebScraping.py (7.8KB)").click()
53             download4 = download4_info.value
54             page.locator("html").click()
55             page.locator("#menu_lecture_material").click()
56             page.get_by_text("LEC 13 Messaging 이대현 조회 118").click()
57             with page.expect_download() as download5_info:
58                 page.get_by_role("link", name="LEC13 - Messaging.pdf (4.6MB)").click()
59                 download5 = download5_info.value
60
61 # -----
62 context.close()
63 browser.close()
64
65
66 with sync_playwright() as playwright:
67     run(playwright)
68
```

At the bottom of the browser window, there is a console log showing the following message:

```
get_by_text("본문내용 바로가기 상단메뉴 바로가기 왼쪽메뉴 바로가기 한국어 English 로그인 이대현 66 20 23 로그아웃 Back to the top ")
```

# 전체 접속 및 다운로드 과정 녹화



The screenshot shows the Playwright Inspector interface. The top bar includes a 'Record' button and a 'Target: Pytest' dropdown. The main area displays a Python test script for logging in and downloading a lecture material. The bottom panel, titled 'Locator Log', shows the locator used for the login button.

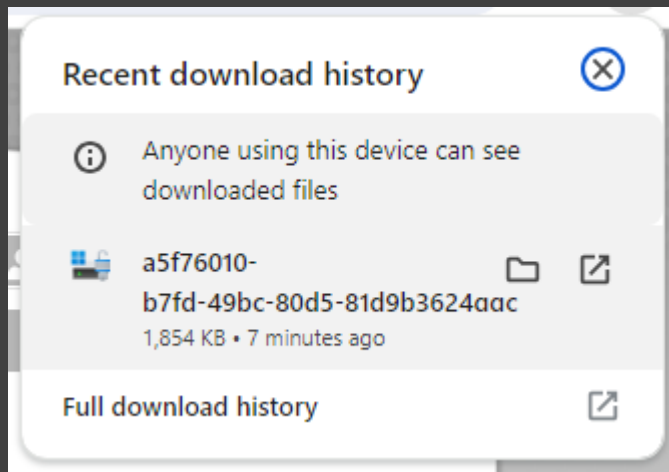
```
1 import re
2 from playwright.sync_api import Page, expect
3
4
5 def test_example(page: Page) -> None:
6     page.goto("https://eclass.tukorea.ac.kr/ilos/main/member/login_form.acl")
7     page.get_by_role("textbox", name="ID").click()
8     page.get_by_role("textbox", name="ID").fill("dustinlee")
9     page.get_by_role("textbox", name="ID").press("Tab")
10    page.get_by_role("textbox", name="Password").fill(" ")
11    page.get_by_role("textbox", name="Password").press("Enter")
12    page.get_by_title("Script Languages", exact=True).click()
13    page.locator("#menu_lecture_material").click()
14    page.get_by_text("LEC 14 Web Scraping (2) (동영상 링크 추가) 이대현 Search").click()
15    with page.expect_download() as download_info:
16        page.get_by_role("link", name="- LEC14 - Web Scraping (2).").click()
17    download = download_info.value
18
```

Locator Log

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

# 파일 다운로드 분석

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





# 파일 다운로드

```
elems = page.locator('a.site-link').all()
for e in elems[: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)')
```

# 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()
```