



ВЕБ-СКРАПІНГ ТА АНАЛІЗ ДАНИХ СОЦІАЛЬНИХ МЕРЕЖ

Робоча програма навчальної дисципліни (Силабус)

1. Реквізити навчальної дисципліни

Рівень вищої освіти	Перший (бакалаврський)
Галузь знань	05 Соціальні та поведінкові науки
Спеціальність	051 Економіка
Освітня програма	Економічна кібернетика
Статус дисципліни	Вибіркова
Форма навчання	Очна (денна)
Рік підготовки, семестр	4 курс, осінній семестр
Обсяг дисципліни	4 кредити ЕКТС /120 годин (лекції: 18 год, практичні: 36 год, СРС: 66 год)
Семестровий контроль/ контрольні заходи	Залік / модульна контрольна робота
Розклад занять	Згідно розкладу: http://roz.kpi.ua 1 лекція (2 години) 1 раз на тиждень; 1 практичне заняття (2 години) 1 раз на тиждень.
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: Стець Олена Вікторівна доцент кафедри економічної кібернетики, канд.фіз.-мат.наук, доцент, робочий кабінет: https://ecocyber.fmm.kpi.ua/uk/stecz-olena-viktorivna/ e-mail: alenka0519@gmail.com Telegram: https://t.me/alenka_stets Комп'ютерні практикуми: Стець Олена Вікторівна доцент кафедри економічної кібернетики, канд.фіз.-мат.наук, доцент, e-mail: alenka0519@gmail.com Telegram: https://t.me/alenka_stets
Розміщення курсу	Платформа дистанційного навчання «Сікорський»: https://classroom.google.com/c/NjQ1MzYyMzU0MTI3 АС «Електронний Кампус» https://campus.kpi.ua Відеолекції та практичні на Youtube channel (за посиланням) Електронний архів наукових та освітніх матеріалів КПІ ім. Ігоря Сікорського (ELAKPI): https://ela.kpi.ua/

2. Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Актуальність вивчення дисципліни «Веб-скрапінг та аналіз даних соціальних мереж» зумовлена інформаційним розвитком суспільства та стрімким розвитком техніки й технологій. В сучасному світі в мережі Інтернеті існує набагато більше інформації, ніж людина може відшукати, проаналізувати, сприйняти та використати за все своє життя. Спростити пошук даних може масштабований спосіб збору інформації, її структуризація та аналіз. Саме для цього потрібно використовувати web-scraping. Отже, актуальним залишається формування у студентів основ інформаційної культури, знань та навичок використання технологій збору та обробки економічних даних та систем програмування під час дослідження соціально-економічних систем та розв'язування завдань фахового спрямування.

Метою дисципліни є формування у студентів системи теоретичних знань і практичних навичок володіння сучасними техніками веб-скрейпінгу, як інструменту професійної діяльності, зокрема для збору інформації.

Предметом дисципліни є методи збору та структуризації даних із мережевих джерел; типи інформаційних систем; принцип роботи пошукових систем; системи керування контентом; концепція web-scraping, парсинг сайтів; технології отримання знань з web.

Програмні компетентності, на формування яких зорієнтована дисципліна:

- ЗК 7 – Навички використання інформаційних і комунікаційних технологій;
- ЗК 8 – Здатність до пошуку, оброблення та аналізу інформації з різних джерел;
- СК 7 – Здатність застосовувати комп'ютерні технології та програмне забезпечення з обробки даних для вирішення економічних завдань, аналізу інформації та підготовки аналітичних звітів.

Програмні результати навчання спрямовані на засвоєння теоретичних знань, розвиток умінь і опанування навичок вирішення складних спеціалізованих задач та практичних проблем економічної сфери:

1. ПРН 19 – Використовувати інформаційні та комунікаційні технології для вирішення соціально-економічних завдань, підготовки та представлення аналітичних звітів;
2. ПРН 23 – Показувати навички самостійної роботи, демонструвати критичне, креативне, самокритичне мислення.
3. ПРН 28 Використовувати методи алгоритмізації та об'єктно-орієнтовані технології для програмування задач в економічному середовищі
4. ПРН 31 – Використовувати різнотипну інформацію, отриману з першоджерел і вторинних джерел використовуючи технологію вебскрайпінгу
5. ПРН 32 – Застосовувати інформаційно-пошукові системи для аналізу та структуризації необхідної інформації аналітичної та науково-дослідної діяльності; застосовувати інформаційно-пошукові системи для аналізу та структуризації необхідної інформації аналітичної та науково-дослідної діяльності

2. Пререквізити та постреквізити дисципліни

Пререквізити: Теоретичною та методологічною базою є дисципліни загально-економічної та професійної підготовки, зокрема «Інформатика Частина 1. Прикладна інформатика в економіці», «Інформатика Частина 2. Основи програмування», «Об'єктно-орієнтоване програмування економічних систем, а також знання основ інформаційно-комунікаційних технологій, володіння методами аналізу, обробки, компонування та подачі даних

Постреквізити: У подальшому дисципліна буде корисною для опанування освітніх компонент «Переддипломна практика» та «Дипломне проектування».

3. 3. Зміст навчальної дисципліни

Перелік тем навчальної дисципліни:

Тема 1. Основи веб-скрейпінгу. Сфери застосування Web scraping.

Тема 2. Етика веб-скрейпінгу. Чи законно парсити дані?

Тема 3. Сучасні інструменти парсингу даних

Тема 4. DOM як основа веб скрейпінгу. Мова запитів XPath.

Тема 5. Python для парсингу даних

Тема 6. Парсинг на Python за допомогою спеціальних бібліотек BeautifulSoup, Scrapy та Selenium.

Тема 7. Скрейпінг соціальних мереж. Огляд API соціальних мереж: Як працюють API Facebook, Twitter, Instagram, LinkedIn.

4. Навчальні матеріали та ресурси

Базова література

1. Мітчелл, Р. (2018). *Web Scraping with Python: Collecting Data from the Modern Web*. 2-ге вид., O'Reilly Media, 416 с.
2. Лутай, І. М. (2020). *Програмування на Python: основи та практика*. Львів: Видавництво "Львівська політехніка". 280 с.
3. Кушнір, А. І. (2021). *Python: алгоритми та структури даних*. Харків: Видавничий дім "Ранок". 250 с.
4. Веб-скрапінг та аналіз даних соціальних мереж : дистанційний курс для здобувачів першого (бакалаврського) рівня вищої освіти спеціальності 051 Економіка / Стець О.В.; КПІ ім. Ігоря Сікорського URL: <https://classroom.google.com/c/NjQ1MzYyMzU0MTI3>

Додаткова література

1. Лінч, Р. (2023). *Web Scraping with Python: A Handy Guide to BeautifulSoup and Selenium*. *Web Dev Byte*. URL: <https://webdevbyte.com/web-scraping-python-beautifulsoup-selenium>
2. Giovanni Romeo. *Elements of Numerical Mathematical Economics with Excel*. 1st Edition. Static and Dynamic Optimization. Academic Press, 2019. P. 816
3. Нестеренко О.В., Ковтунець О.В., Фаловський О.О. *Інтелектуальні системи і технології*. Ввідний курс: Навч. посібник. К.: Національна академія управління, 2017. 99 с. URL: <http://e.ieu.edu.ua/handle/123456789/646>

Інформаційні ресурси

1. *Programming for Everybody (Getting Started with Python)*. URL: [Доступно з: https://www.coursera.org/learn/python](https://www.coursera.org/learn/python)

На лекції та практичні роботи є відеоматеріал, який викладено на Youtube channel та доступний студентам за посиланням.

Дисципліна	Веб-скрапінг та аналіз даних соціальних мереж
Вимоги до початку вивчення	Теоретичною та методологічною базою є дисципліни загальноекономічної та професійної підготовки, зокрема «Інформатика», «Інформаційні системи та технології в управлінні», «Програмне забезпечення моделювання економічних систем: Основи програмування та алгоритмічні мови», «Об'єктно-орієнтоване програмування економічних систем», «Веб програмування», а також знання основ інформаційно-комунікаційних технологій, володіння методами аналізу, обробки, компонування та подачі даних
Що буде вивчатися	Аналітика в Інтернеті; методологія збору та структуризації даних із мережевих джерел, зокрема соцмереж; типи інформаційних систем; принцип роботи пошукових систем; системи керування контентом; концепція бізнес-аналітичного рішення; концепція web-scraping, парсинг сайтів; класифікація способів вилучення

	інформації з веб-джерел; моделі інформаційного пошуку; технології отримання знань з web
Чому це цікаво/треба вивчати	В сучасному світі в мережі Інтернеті існує набагато більше інформації, ніж людина може відшукати, проаналізувати, сприйняти та використати за все своє життя. Спростити пошук даних може масштабований спосіб збору інформації, її структуризація та аналіз. Саме для цього потрібно використовувати web-scraping.
Чому можна навчитися (результати навчання)	<ul style="list-style-type: none"> - Застосовувати інформаційно-пошукові системи для аналізу та структуризації необхідної інформації аналітичної та науково-дослідної діяльності; застосовувати інформаційно-пошукові системи для аналізу та структуризації необхідної інформації аналітичної та науково-дослідної діяльності - Застосовувати контент-аналіз з використанням пошукових веб-браузерів - Використовувати в проєктованих та експлуатованих інформаційних системах та технологіях сучасні засоби Інтернет-програмування - Ознайомитись з методами та сценаріями організації даних в соцмережах; аналіз та монетизація відповідних даних - Застосовувати сучасні інформаційні технології та спеціалізоване програмне забезпечення у соціально-економічних дослідженнях та в управлінні соціально-економічними системами
Як можна користуватися набутими знаннями і уміннями (компетентності)	<ul style="list-style-type: none"> - Здатність аналізувати та виявляти застарілу та недостовірної інформацію - Здатність формувати у студентів уявлення про можливості, техніку та межі новітніх кількісних та якісних методів аналізу економічної інформації - Формування у студентів цілісного уявлення про основні принципи функціонування соціальних мереж; аналіз та монетизація відповідних даних - Здатність використовувати сучасні інформаційні технології, методи та прийоми дослідження економічних та соціальних процесів, адекватні встановленим потребам дослідження

4. 6. Самостійна робота здобувача вищої освіти

Навчальним планом передбачено 66 годин самостійної роботи. На самостійну роботу вносятся: підготовка до аудиторних занять, до виконання модульної контрольної роботи, підготовка до складання семестрового контролю.

Вид самостійної роботи, обсяг годин на виконання

№ з/п	Обсяг годин	Вид самостійної роботи
1	2	3
1	56	Підготовка до аудиторних занять
2	4	Підготовка до виконання модульної контрольної роботи
3	6	Підготовка до складання семестрового контролю у формі заліку
Разом	66	

5. Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Правила відвідування занять. У режимі очного навчання заняття відбуваються в аудиторії згідно розкладу занять, у режимі дистанційного навчання заняття відбуваються у вигляді онлайн-конференції у програмі Zoom - посилання на конференцію видається на початку семестру. Відвідування занять є вільним, бали за присутність на лекції не додаються, і штрафні бали за пропуски занять не передбачено. Втім, вагома частина рейтингу студента формується через активну участь у заходах на практичних заняттях, а саме у вирішенні завдань, груповій та індивідуальній роботі. Тому пропуск практичного заняття не дає можливість отримати студенту бали у семестровий рейтинг. На заняттях студенту дозволяється користуватись інтерактивними засобами навчання, в т.ч. виходити в Інтернет із метою пошуку навчальної або довідкової інформації, якщо це передбачено тематикою завдання. Активність студента на парах, його готовність до дискусій та участь в обговоренні навчальних питань може бути оцінена заохочувальними балами на розсуд викладача

Дистанційний режим навчання. У разі запровадження обмежень на відвідування університету у разі організації освітнього процесу у змішаному /дистанційному пов'язаних з введенням режиму воєнного стану в державі (або карантину), освітній процес здійснюється відповідно до Положення про дистанційне навчання в КПІ ім. Ігоря Сікорського (<https://osvita.kpi.ua/index.php/node/188>), У режимі дистанційного навчання заняття відбуваються у вигляді онлайн-конференції на платформі Zoom. Посилання на конференцію видається на початку семестру і розміщується в АС «Електронний кампус». З метою забезпечення якісної підготовки здобувачів, дистанційний курс дисципліни розміщено на Платформа дистанційного навчання «Сікорський» (<https://classroom.google.com/c/NjQ1MzYxNTY2Mzkz?cjc>). Результати оцінювання висвітлюють у АС «Електронний кампус» на особистій сторінці здобувача (<https://ecampus.kpi.ua>).

Правила поведінки на заняттях. Дотримання норм етичної поведінки визначених у Кодексі честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» <https://kpi.ua/code> На території університету студенти мають поводити себе відповідно до Правил внутрішнього розпорядку <https://kpi.ua/admin-rule> В аудиторіях на лекційних та практичних заняттях користуватись мобільним телефоном потрібно у беззвучному режимі і тільки для пошуку необхідної для виконання завдань інформації, у т.ч. в Інтернеті. Під час дистанційного режиму навчання бажано мати та використовувати інформацію розміщену на Платформі дистанційного навчання «Сікорський».

Позааудиторні заняття та залучення професіоналів-практиків. Під час вивчення дисципліни можливі позааудиторні заняття, що включають відвідування міжнародних конференцій та інших науково-практичних заходів в межах тематики дисципліни за умови активної участі у таких заходах. Для опанування і поглиблення практичних навичок на заняття можуть бути запрошені професіонали-практики (стейкхолдери) за попереднім узгодженням.

Визнання результатів навчання, набутих у неформальній/інформальній освіті. Положення про визнання результатів навчання, набутих у неформальній / інформальній освіті (<https://osvita.kpi.ua/index.php/node/179>) регламентує визнання результатів навчання, набутих у неформальній/інформальній освіті. У разі проходження здобувачем онлайн курсу чи іншого елемента неформальної освіти за наведеною у розділі «Додаткова інформація з дисципліни» здобувачеві/здобувачці можуть бути зараховані окремі змістовні модулі або теми дисципліни. В такому разі здобувач звільняється від виконання відповідних завдань, отримуючи за них максимальний бал відповідно до рейтингової системи оцінювання. У разі навчання на неформальній освіті за самостійного її обрання проходиться процедура валідації, що передбачає подання здобувачем заяви на ім'я декана, декларації підтверджувальних документів. Рішення про визнання чи не визнання приймається комісією у складі завідувача кафедри, викладача, гаранта освітньо-професійної програми.

Правила призначення заохочувальних та штрафних балів. Ключовими заходами при викладанні дисципліни є ті, які формують семестровий рейтинг студента. Тому студенти мають своєчасно виконувати завдання на практичних заняттях, писати модульну контрольну роботу. Заохочувальні бали студент може отримати за поглиблене вивчення окремих тем курсу, що може бути

представлене у вигляді наукових тез, наукової статті, додаткових завдань тощо, а також за активну участь у дискусіях на практичних та лекційних заняттях. Штрафні бали не передбачаються.

Політика оцінювання контрольних заходів. Оцінювання контрольних заходів відбувається відповідно до Положення про систему оцінювання результатів навчання в КПІ ім. Ігоря Сікорського <https://osvita.kpi.ua/node/37>, Положення про поточний, календарний та семестровий контроль результатів навчання в КПІ ім. Ігоря Сікорського <https://osvita.kpi.ua/index.php/node/32> Нижня межа позитивного оцінювання кожного контрольного заходу складає не менше 60% від балів, визначених для цього контрольного заходу, а негативний результат оцінюється в 0 балів.

Поточний контроль проводиться згідно рейтингової системи оцінювання.

Календарний контроль проводиться двічі на семестр, а також передбачено проведення модульної контрольної роботи, яка здійснюється у вигляді письмової контрольної роботи.

Умовою отримання позитивної оцінки з календарного контролю з навчальної дисципліни є значення поточного рейтингу здобувача не менше, ніж 50% від максимально можливого на час проведення такого контролю. Семестровий контроль проводиться у формі заліку за РСО-1 («жорстка» РСО), що передбачає оцінювання виконаних завдань впродовж семестру, а під час заліку попередній рейтинг здобувача (за винятком балів за семестрове індивідуальне завдання) скасовується і студент отримує оцінку з урахуванням результатів залікової контрольної роботи. Цей варіант формує відповідальне ставлення здобувача до прийняття рішення про виконання залікової контрольної роботи, змушує його критично оцінити рівень своєї підготовки та ретельно готуватися до заліку.

Результати оцінювання висвітлюються у АС «Електронний кампус» на особистій сторінці здобувача (<https://ecampus.kpi.ua>).

Політика дедлайнів та перескладань. Формування семестрового рейтингу студента ґрунтується своєчасному виконанні поточних завдань, написанні модульної контрольної роботи згідно графіку викладання дисципліни. Якщо контрольні заходи, або виконання завдань пропущені з поважних причин (хвороба або вагомі життєві обставини), надається можливість додатково скласти завдання протягом найближчого тижня або відпрацювати пропущення заняття шляхом виконання індивідуальних завдань. Детальніше згідно Положення про поточний, календарний та семестровий контроль результатів навчання в КПІ ім. Ігоря Сікорського <https://kpi.ua/files/n3277.pdf> Порядок ліквідації академічної заборгованості та перескладання семестрового контролю регулюється Положенням про поточний, календарний та семестровий контроль результатів навчання в КПІ ім. Ігоря Сікорського <https://osvita.kpi.ua/index.php/node/32> Здобувач, у якого за результатами семестрового контролю виникла академічна заборгованість, має право її ліквідувати відповідно до Положення про надання додаткових освітніх послуг здобувачам вищої освіти в КПІ ім. Ігоря Сікорського <https://osvita.kpi.ua/index.php/node/177>

Процедура оскарження результатів контрольних заходів. У випадку не згоди здобувача з оцінкою за результатами контрольного заходу, він має право подати апеляцію у день оголошення результатів відповідного контролю на ім'я декана факультету за процедурою визначеною Положенням про апеляції в КПІ ім. Ігоря Сікорського <https://osvita.kpi.ua/index.php/node/182>

Політика щодо академічної доброчесності. Необхідним під час виконання завдань з дисципліни є дотримання політика та принципів академічної доброчесності, які, у тому числі викладено у Кодексі честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» <https://kpi.ua/code>, Положенні про систему запобігання академічному плагіату в КПІ ім. Ігоря Сікорського <https://osvita.kpi.ua/node/47>.

Політика використання штучного інтелекту. Використання штучного інтелекту регламентується «Політикою використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського» (<https://osvita.kpi.ua/node/1225>). Усі завдання, як під час виконання навчальних завдань з дисципліни, так і індивідуальні завдання, мають бути результатом власної оригінальної роботи здобувача. Використання ШІ для автоматичної генерації відповідей без подальшого їх аналізу та доопрацювання заборонено. Здобувачам не рекомендується покладатися на ШІ як на єдине джерело інформації. Важливо перевіряти та аналізувати отримані дані з інших авторитетних

джерел. Усі випадки використання ШІ для виконання завдань мають бути чітко вказані та задокументовані. Це стосується як використання текстових генераторів, так і інших інструментів ШІ. Використання ШІ має відповідати принципам академічної доброчесності. Недотримання цього положення розглядатиметься як порушення академічної етики.

6. 8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Оцінювання передбачає застосування рейтингової системи другого типу згідно Положення про систему оцінювання результатів навчання в КПІ ім. Ігоря Сікорського (PCO-1) (<https://osvita.kpi.ua/node/37>). Максимальний сумарний рейтинг за курс – 100 балів. Нижня межа позитивного оцінювання кожного контрольного заходу складає 60% від його рейтингового балу.

Контрольні заходи та критерії їх оцінювання

8.1. Поточний контроль (тах 100 балів)

1. Опитування (тах 10 балів)

1. *Відповіді на практичних заняттях (10 експрес-опитувань):* опитування (відповіді на запитання з теми лекції або по деталізації виконання практичної роботи). *Розрахунок за одну відповідь:*

- | | |
|---------|--|
| 1 бал | 6. активна робота, повні відповіді під час опитування робота з доповненням до відповідей інших здобувачів; |
| 0 балів | 7. пасивність |

2. Виконання комп'ютерних практикумів **(10 КП тах 80 бали)**. *Розрахунок за одне завдання:*

- | | |
|-----------|--|
| 10 балів | – КП виконано у повному обсязі з належним аналізом та обґрунтуванням отриманих даних; |
| 8-9 балів | – КП виконано у повному обсязі з незначними помилками, які суттєво не впливають на її результат; |
| 7 балів | – КП виконано з помилками, які мають вплив на її результат, обґрунтування та висновки; |
| 6 балів | – КП виконано не у повному обсязі, є грубі помилки у розрахунках, отримані дані не обґрунтовано; |
| 0 балів | – КП не виконано |

2. Модульна контрольна робота (тах 10 балів)

Модульна контрольна робота складається з тестових завдань загальною сумою 10 балів

8.2. Календарний контроль.

Календарний контроль проводиться двічі на семестр.

- | | |
|--------------|---|
| 7 тиждень – | Умова отримання атестації: поточний рейтинг не менше 15 балів, виконано не менше ніж 30 % з обсягу завдань; |
| 14 тиждень - | Умова отримання атестації: поточний рейтинг не менше 50 балів, виконано не менше ніж шість КП |

8.3. Семестровий контроль (залік) (тах 100 балів)

Складається з поточного рейтингу за виконання всіх видів робіт

Студенти, які виконали всі умови допуску до заліку та мають рейтингову оцінку 60 і більше балів, отримують відповідну до набраного рейтингу оцінку без додаткових випробувань. Сума рейтингових балів, отриманих студентом протягом семестру, переводиться до підсумкової оцінки згідно з таблицею.

Якщо сума балів менша за 60, але виконані і зараховані МКР, студент виконує залікову контрольну роботу. У цьому разі підсумкова оцінка отримується за суму балів, які студент отримав за залікову контрольну роботу, що переводиться до підсумкової оцінки згідно з таблицею.

Студент, який у семестрі отримав більше 60 балів, але бажає підвищити свій результат, може взяти участь у заліковій контрольній роботі. У цьому разі підсумкова оцінка отримується за суму балів, які студент отримав за залікову контрольну роботу, що переводиться до підсумкової оцінки згідно з таблицею.

Розрахунок балів за виконання завдань залікової контрольної роботи:

- | | |
|-------------|--|
| – | теоретичні питання (3 питання), за одне питання: |
| 19-20 балів | – повна відповідь (не менше 90% потрібної інформації); |

- 17-18 балів – достатньо повна відповідь (не менше 75% потрібної інформації) або повна відповідь з незначними неточностями;
 - 15-16 балів – неповна відповідь (не менше 60% потрібної інформації) та незначні помилки;
 - 12 балів – неповна відповідь (не менше 60% потрібної інформації) та незначні помилки;
 - 0 балів – немає відповіді.
 - аналітичне завдання. Передбачає виконання одного завдання:
 - 40-38 балів – повне виконання завдання (не менше 95%);
 - 37-30 балів – достатньо повне виконання завдання (не менше 75%) або повне виконання з незначними неточностями;
 - 29-24 балів – неповне виконання завдання (не менше 60%) та незначні помилки;
 - 0 балів – завдання не виконано.
- Максимальний бал за курс – 100 балів.

Відповідність рейтингових балів оцінкам за університетською шкалою:

Кількість балів	Оцінка
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

7.

9. Додаткова інформація з дисципліни (освітнього компонента)

Перелік питань, які виносяться на семестровий контроль (додаток А до силабусу)

Можливе зарахування сертифікатів проходження дистанційних чи онлайн курсів за відповідною тематикою у рейтинг здобувача. Рекомендовані онлайн курси: «Програмування для всіх: основи Python» (платформа онлайн освіти Prometheus); Programming for Everybody (Getting Started with Python) (платформа онлайн освіти Coursera)

У навчальному процесі використовується програмні продукти: NotePad++ (безкоштовно (freeware)); MS Word та MS Excel (Підписка на продукт Microsoft 365), Google Workspace for Education (безкоштовно (freeware)).

Викладення дисципліни може бути переведено у дистанційну форму за відповідних умов згідно розпоряджень університету.

Робочу програму навчальної дисципліни (силабус):

Складено доцент, канд. фіз.-мат. наук

Стець Олена Вікторівна

Ухвалено кафедрою економічної кібернетики (протокол № 18 від 18.06.2024р.)

Погоджено Методичною комісією факультету менеджменту та маркетингу (протокол № 10 від 24.06.2024 р.)

ДОДАТОК А

Перелік питань для підготовки до семестрового контролю

8. Що таке веб-скрейпінг і як він працює?
9. У яких сферах застосовують веб-скрейпінг?
10. Які є переваги та недоліки веб-скрейпінгу порівняно з іншими методами збору даних?
11. Які існують типи веб-скрейпінгу (статичний та динамічний)?
12. Що таке CAPTCHA і як вона впливає на процес веб-скрейпінгу?
13. Які основні етичні проблеми виникають при веб-скрейпінгу?
14. Які міжнародні правові акти регулюють збір даних через веб-скрейпінг?
15. Які є методи уникнення порушення правил конфіденційності під час скрейпінгу?
16. Що таке "Terms of Service" і як вони стосуються веб-скрейпінгу?
17. Чи можна вважати скрейпінг публічно доступних даних законним? Чому?
18. Які наслідки може мати порушення правил веб-скрейпінгу на конкретних платформах?
19. Як різні країни регулюють використання веб-скрейпінгу?
20. Які найпопулярніші інструменти для веб-скрейпінгу на сьогодні?
21. Що таке DOM і як він використовується при веб-скрейпінгу?
22. Як взаємодіяти з елементами DOM для збору даних?
23. Що таке XPath і яку роль він відіграє у веб-скрейпінгу?
24. Як використовувати XPath для витягування даних зі складних HTML-структур?
25. Які є основні відмінності між CSS-селекторами та XPath при парсингу даних?
26. У яких випадках XPath є ефективнішим за інші методи пошуку елементів у DOM?
27. Які виклики виникають при використанні XPath для динамічних сторінок?
28. Чому Python є популярною мовою для веб-скрейпінгу?
29. Які є основні бібліотеки Python для парсингу веб-сайтів?
30. Які типи даних найчастіше використовують при зберіганні результатів парсингу у Python?
31. Як використовувати Python для відправки HTTP-запитів і обробки відповідей?
32. У чому різниця між BeautifulSoup, Scrapy та Selenium?
33. Які переваги дає використання Selenium для динамічних веб-сторінок?
34. Як можна використовувати API для збору даних, замість веб-скрейпінгу?
35. Які переваги надають готові веб-скрейпінгові сервіси, порівняно з власноруч написаними скриптами?
36. Що таке headless-браузер, і як його можна використовувати для скрейпінгу?
37. Як обрати інструмент для парсингу залежно від типу сайту?
38. Як за допомогою BeautifulSoup можна парсити статичні сторінки?
39. Як працює Scrapy і для яких завдань він найкраще підходить?
40. Як за допомогою Selenium можна взаємодіяти з веб-сторінками та автоматизувати браузерні дії?
41. Як можна поєднати використання BeautifulSoup та Selenium у парсингу?
42. Як працюють API соціальних мереж, таких як Facebook, Twitter, Instagram та LinkedIn?