

СИЛЛАБУС
Осенний семестр 2025 -2026 учебного года
Образовательная программа «БВ06102 – Информационные системы»

ID и наименование дисциплины	Самостоятельная работа обучающегося (СРС)	Кол-во кредитов			Общее кол-во кредитов	Самостоятельная работа обучающегося под руководством преподавателя (СРСР)
		Лекции (Л)	Практ. занятия (ПЗ)	Лаб. занятия (ЛЗ)		
86193 Поиск и извлечение данных	4	1.5	1.5	3	6	7
АКАДЕМИЧЕСКАЯ ИНФОРМАЦИЯ О ДИСЦИПЛИНЕ						
Формат обучения	Цикл, модуль компонент	Типы лекций	Типы практических занятий	Форма и платформа итогового контроля		
Офлайн	П, КпВ	Проблемно-ориентированный	Изучение концепций поиска данных и веб-скрапинга	Устный офлайн		
Лектор - (ы)	Карюкин В.И.					
e-mail:	vladislav.karyukin@kaznu.kz					
Телефон:	+77019405992					
Ассистент- (ы)	Карібаева А.С.					
e-mail:	karibayeva.aidana@kaznu.kz					
Телефон:	+77772322091					
АКАДЕМИЧЕСКАЯ ПРЕЗЕНТАЦИЯ ДИСЦИПЛИНЫ						
Цель дисциплины	Ожидаемые результаты обучения (РО)*			Индикаторы достижения РО (ИД)		
Целью дисциплины является формирование способностей применять практические навыки веб-поиска и ранжирования данных. Дисциплина направлена на изучение базовых понятий информационного поиска и основных принципов информационно-поисковых систем, обработку и анализ данных различного объема, анализ информации из различных источников и баз данных с использованием информационных, компьютерных и сетевых технологий.	1. Когнитивный. Понимать теоретические основы поиска и индексации данных, знать основные алгоритмы и модели			1.1 Объясняет основные модели поиска и приводит примеры их применения		
				1.2 Классифицирует методы поиска и указывает их сильные и слабые стороны		
	2. Функциональный. Использовать библиотеки Beautiful Soup, Requests, Selenium Web Driver для реализации веб-скрапинга			2.1 Выполняет установку и настройку библиотек для веб-скрапинга		
				2.2 Реализует настройку поискового движка и формирует поисковые запросы		
	3. Функциональный. Разрабатывать прототипы систем извлечения информации, включая парсинг данных, построение индекса и реализацию пользовательских запросов			3.1 Умеет собирать данные из различных источников (веб-страницы, API) для дальнейшего поиска		
			3.2 Демонстрирует владение инструментами для предварительной обработки данных			
4. Системный. Внедрять модули поиска и веб-скрапинга, обеспечивающие комплексную работу с информационными потоками			4.1 Разрабатывает модуль поиска и подключает его к веб-приложению или сервису			
			4.2 Демонстрирует работу интегрированной системы на тестовом наборе данных, подтверждая корректность поиска и извлечения данных			
5. Когнитивный. Анализировать релевантность полученной информации и оценивать эффективность информационно-поисковых систем			5.1 Описывает архитектуру современных поисковых систем (веб-скрапинг, индексация, ранжирование)			
			5.2 Формулирует выводы по результатам экспериментов и интерпретирует статистические показатели			
Пререквизиты	101286 - Web-программирование, 101817 - Программирование на языке Python					
Постреквизиты	101293 - Интеллектуальный анализ данных					
Учебные ресурсы	Литература:					

Основная

1. Allen B. Downey. *Data Structures and Information Retrieval in Python*. 2021.
2. Jimmy Lin et al. "Pyserini: An Easy-to-Use Python Toolkit to Support Replicable IR Research with Sparse and Dense Representations." SIGIR / arXiv, 2021.
3. Craig Macdonald & Nicola Tonellotto. "Declarative Experimentation in Information Retrieval using PyTerrier." 2020.
4. Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Jingfang Xu, Xueqi Cheng. "DeepRank: A New Deep Architecture for Relevance Ranking in Information Retrieval." 2017.
5. Python for Everybody: Exploring Data in Python 3 by Dr. Charles Russell Severance, Sue Blumenberg, Elliott Hauser, Aimee Andrion, 2016.

Дополнительная

1. Natural Language Processing with Python and Spacy: A Practical Introduction, Yuli Vasiliev, 2021
2. Learning Scientific Programming with Python, Christian Hill, 2021

Исследовательская инфраструктура

1. Бизнес-инкубатор №12
2. Лабораторная аудитория 514
3. Лабораторная аудитория 323

Профессиональные научные базы данных

1. Scopus – <https://www.scopus.com>
2. Elsevier – <https://www.elsevier.com/>
3. Researchgate – <https://www.researchgate.net/>

Интернет-ресурсы

1. <https://www.udemy.com/course/web-scraping-course-in-python-bs4-selenium-and-scrapy/>
2. <https://www.udemy.com/course/introduction-to-data-extractionweb-scraping-in-python/>
3. <https://www.udemy.com/course/master-python-web-scraping-automation-using-bs4-selenium/>

Программное обеспечение

1. Sublime
2. Visual code
3. PyCharm
4. Anaconda Navigator Python
5. Microsoft Visual Studio

Академическая политика дисциплины

Академическая политика дисциплины определяется Академической политикой и Политикой академической честности КазНУ имени аль-Фараби.

Документы доступны на главной странице ИС Univer.

Интеграция науки и образования. Научно-исследовательская работа студентов, магистрантов и докторантов – это углубление учебного процесса. Она организуется непосредственно на кафедрах, в лабораториях, научных и проектных подразделениях университета, в студенческих научно-технических объединениях. Самостоятельная работа обучающихся на всех уровнях образования направлена на развитие исследовательских навыков и компетенций на основе получения нового знания с применением современных научно-исследовательских и информационных технологий. Преподаватель исследовательского университета интегрирует результаты научной деятельности в тематику лекций и семинарских (практических) занятий, лабораторных занятий и в задания СРОП, СРО, которые отражаются в syllabusе и отвечают за актуальность тематик учебных занятий и заданий.

Посещаемость. Дедлайн каждого задания указан в календаре (графике) реализации содержания дисциплины. Несоблюдение дедлайнов приводит к потере баллов.

Академическая честность. Практические/лабораторные занятия, СРО развивают у обучающегося самостоятельность, критическое мышление, креативность. Недопустимы плагиат, подлог, использование шпаргалок, списывание на всех этапах выполнения заданий.

Соблюдение академической честности в период теоретического обучения и на экзаменах помимо основных политик регламентируют «Правила проведения итогового контроля», «Инструкции для проведения итогового контроля осеннего/весеннего семестра текущего учебного года», «Положение о проверке текстовых документов обучающихся на наличие заимствований».

Документы доступны на главной странице ИС Univer.

Основные принципы инклюзивного образования. Образовательная среда университета задумана как безопасное место, где всегда присутствуют поддержка и равное отношение со стороны преподавателя ко всем обучающимся и обучающимся друг к другу независимо от гендерной, расовой/ этнической принадлежности, религиозных убеждений, социально-экономического статуса, физического здоровья студента и др. Все люди нуждаются в поддержке и дружбе ровесников и сокурсников. Для всех студентов достижение прогресса скорее в том, что они могут делать, чем в том, что не могут. Разнообразие усиливает все стороны жизни.

6	Л 6. Продвинутые техники работы с BeautifulSoup: использование CSS-селекторов, очистка и нормализация текста, парсинг вложенных структур	1	1
	ПЗ 6. Использование CSS-селекторов и XPath для поиска элементов страницы	1	5
	ЛЗ 6. Разработка системы сбора данных с использованием Selenium	2	5
	СРСП 2. Консультации по выполнению СРС 2 на тему «Создание скрапера для сбора данных их социальных сетей»		
7	Л 7. Основы работы с Selenium WebDriver: автоматизация браузера, взаимодействие с динамическим контентом и обработка элементов страницы	1	1
	ПЗ 7. Парсинг данных с веб-сайтов с использованием Selenium WebDriver	1	5
	ЛЗ 7. Сравнительный анализ эффективности Selenium и requests_html при работе с динамическими страницами	2	5
	СРСП 3. Прием СРС 2 в форме реферата		30
8	Л 8. Расширенные возможности Selenium: работа с формами, кнопками, скроллингом, JavaScript-элементами и снятие скриншотов	1	1
	ПЗ 8. Эмуляция нажатия кнопок в Selenium WebDriver	1	5
	ЛЗ 8. Разработка парсера для извлечения данных из сайта, имеющего сложную структуру	2	5
			100
Рубежный контроль 1			
9	Л 9. Извлечение данных через API как альтернатива скрапину: основы REST API, работа с JSON и практические примеры запросов	1	1
	ПЗ 9. Использование API для получения данных с веб-страниц	1	5
	ЛЗ 9. Разработка приложения для скрапинга с использованием API	2	5
	СРСП 4. Консультации по выполнению СРС 3 на тему «Создание приложения для парсинга данных с YouTube»		
10	Л 10. Хранение и структурирование собранных данных: сохранение в CSV, JSON и базы данных, интеграция с библиотекой pandas	1	1
	ПЗ 10. Веб-скрапинг новостей с сайтов, имеющих постраничную навигацию	1	5
	ЛЗ 10. Реализация приложения по веб-скрапину новостей с нескольких страниц сайта с постраничной навигацией	2	5
МОДУЛЬ 3 Разработка приложения веб-скрапинга			
11	Л 11. Фреймворк Scrapy: основы архитектуры, создание пауков, конфигурация проекта и работа с пайплайнами	1	1
	ПЗ 11. Использование библиотеки Scrapy для парсинга веб-страниц	1	5
	ЛЗ 11. Разработка программы веб-скрапинга с использованием Scrapy и сохранением данных в базу	2	5
	СРСП 5. Прием СРС 3 в форме реферата		10
12	Л12. Продвинутые возможности Scrapy: middlewares, работа с куками и сессиями, обработка больших объемов данных	1	1
	ПЗ 12. Применение продвинутых возможностей Scrapy для парсинга сайтов	1	5
	ЛЗ 12. Реализация работы программы со Scrapy	2	5
	СРСП 6. Консультации по выполнению СРС 4 на тему «Разработка комплексного веб-приложения по парсингу данных из различных интернет источников»		
13	Л 13. Парсинг JavaScript сайтов с использованием Scrapy	1	1
	ПЗ 13. Парсинг динамических данных, генерируемых JavaScript	1	5
	ЛЗ 13. Создание программы для парсинга динамических данных веб-сайтов	2	5
14	Л 14. Парсинг данных с использованием ChatGPT	1	1
	ПЗ 14. Применение ChatGPT для парсинга данных	1	5
	ЛЗ 14. Разработка приложения для сбора и анализа мультимедийного контента: изображения, видео, документы	2	5
	СРСП 7. Прием СРС 4 в форме реферата		13
15	Л 15. Разработка веб-приложения по парсингу данных	1	1
	ПЗ 15. Создание многофункционального веб-приложения парсинга данных	1	5
	ЛЗ 15. Разработка системы с автоматическим запуском для сбора данных из порталов и социальных сетей	2	5
Рубежный контроль 2			100
Итоговый контроль (экзамен)			100
ИТОГО за дисциплину			100

РУБРИКАТОР СУММАТИВНОГО ОЦЕНИВАНИЯ
КРИТЕРИИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

СРС 1. Создание веб-скрапера для нескольких новостных порталов (25% от 100% РК1)

Критерий	«Отлично» 21-25%	«Хорошо» 11-20%	«Удовлетворительно» 6-10%	«Неудовлетворительно» 0-5%
Знание и понимание основных элементов веб-скрапера для нескольких новостных порталов	Отличное понимание степени соответствия, актуальности и достоверности найденных данных. Знание и понимание всех основных элементов и операций веб-скрапера для нескольких новостных порталов	Хорошее понимание степени соответствия, актуальности и достоверности найденных данных. Знание больше части операций веб-скрапера для нескольких новостных порталов	Ограниченное понимание степени соответствия, актуальности и достоверности элементов и операций веб-скрапера для нескольких новостных порталов	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности найденных данных. Отсутствие знания элементов и операций веб-скрапера для нескольких новостных порталов
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических и синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и корректность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать за содержанием. Много ошибок в тексте

СРС 2. Создание скрапера для сбора данных их социальных сетей (30% от 100% РК1)

Критерий	«Отлично» 21-30%	«Хорошо» 11-20%	«Удовлетворительно» 6-10%	«Неудовлетворительно» 0-5%
Работа со скрапером для сбора данных их социальных сетей	Отличное понимание степени соответствия, актуальности и достоверности работы со скрапером для сбора данных их социальных сетей	Хорошее понимание степени соответствия, актуальности и достоверности найденных данных. Знание больше части операций работы со скрапером для сбора данных их социальных сетей	Ограниченное понимание степени соответствия, актуальности и достоверности операций по работе со скрапером для сбора данных их социальных сетей	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности работы со скрапером для сбора данных их социальных сетей
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических и синтаксических ошибок в	Отсутствие программного кода или наличие нескольких строк кода

			программном коде, что делают его практически неработоспособным	
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и корректность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать за содержанием. Много ошибок в тексте

СРС 3. Создание приложения для парсинга данных с YouTube (10% от 100% ПК2)

Критерий	«Отлично» 8-10%	«Хорошо» 5-7%	«Удовлетворительно» 3-4%	«Неудовлетворительно» 0-2%
Работа с парсингом данных с YouTube	Понимание степени соответствия, актуальности и достоверности работы с парсингом данных с YouTube	Понимание степени соответствия, актуальности и достоверности найденных данных. Знание больше части операций создания Django приложения	Ограниченное понимание степени соответствия, актуальности и достоверности операций по созданию Django приложения	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности работы с базами данных. Отсутствие знания операций создания Django приложения
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических и синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и корректность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать за содержанием. Много ошибок в тексте


СРС 4. Разработка комплексного веб-приложения по парсингу данных из различных интернет-источников (13% от 100% ПК2)


Критерий	«Отлично» 10-13%	«Хорошо» 7-9%	«Удовлетворительно» 3-6%	«Неудовлетворительно» 0-2%
Знание и понимание основных элементов машинного и глубокого обучения в Django	Отличное понимание степени соответствия, актуальности и достоверности найденных данных. Знание и понимание всех основных элементов и операций машинного и глубокого обучения в Django	Хорошее понимание степени соответствия, актуальности и достоверности найденных данных. Знание больше части машинного и глубокого обучения в Django	Ограниченное понимание степени соответствия, актуальности и достоверности машинного и глубокого обучения в Django	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности найденных данных. Отсутствие знания элементов и операций машинного и глубокого обучения в Django
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических и синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода

Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и корректность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать за содержанием. Много ошибок в тексте
------------------	--	---	---	---

Декан  Иманкулов Т.С.

Председатель Академического комитета
по качеству преподавания и обучения  Ж.А. Бурибаев

Заведующий кафедрой  Шормакова А.Н.

Лектор  Карюкин В.И.

