

Дисципліна	Основи науки про дані
Кафедра	Обчислювальної техніки
Рівень ВО	Перший (бакалаврський)
Можливі обмеження	Кількість студентів на курс від 60 до 150
Курс, семестр	3 курс, 5 семестр
Обсяг	4 кредити, 120 годин
Мова викладання	Українська
Вимоги до початку вивчення	Базові знання з вищої математики (математичний аналіз, лінійна алгебра та аналітична геометрія), теорія ймовірностей та математична статистика, програмування
Що буде вивчатися	Загальні поняття про сутність аналізу даних, застосування принципу Бонферроні, методу пошуку найближчих сусідів, шинглінгу документів. Принципи збереження схожості скорочених наборів, локально-чутливе хешування та його застосування, міри відстані. Поняття про аналіз потоків даних різної природи та складності та аналіз посилань. Модель ринкової корзини та алгоритм A-Priori
Чому це цікаво/треба вивчати	Наука про дані має вагомe практичне значення для вирішення актуальних повсякденних проблем. Застосування науки про дані змінює спосіб життя, праці й мислення, оскільки однією з умов успішного розвитку світової економіки на сучасному етапі є можливість фіксувати й аналізувати накопичені набори даних. Сучасні ефективні методи роботи з даними сприяють як підвищенню конкурентоспроможності різних сфер економічної діяльності, так і фахівців, які володіють цими методами.
Чому можна навчитися (результати навчання)	Створювати моделі для даних з застосуванням узагальнення та виділення ознак. Використовувати методи обчислення жакартової подібності множин. Застосовувати міри відстані для визначення подібності, зокрема, косинусну відстань, відстань редагування, відстань Хеммінга, локально-чутливі функції. Аналізувати потоки даних шляхом побудови моделі поточкових даних та фільтрації потоків, здійснювати підрахунок різних елементів у потоці, зокрема, підрахунок у вікні та згасаючих вікнах, використовувати алгоритми DGIM. Здійснювати аналіз посилань з застосуванням PageRank, його ефективне обчислення, обчислювати тематично чутливий PageRank, визначати спам посилань.
Як можна користуватися набутими знаннями і	для формування початкових наборів даних з застосуванням принципу Бонферроні;

уміннями (компетентності)	• для пошуку текстуально подібних документів (плагіату, дзеркальних сторінок, статей з одного джерела); • для визначення подібності он-лайнних покупок та рейтингу фільмів; • для розпізнавання об'єктів, співставлення відбитків пальців та газетних статей; • для створення програмних систем ефективної фільтрації небажаних елементів потоку (вхідні потоки електронної пошти); • для створення телепорт-наборів.
Інформаційне забезпечення	Силабус дисципліни, навчально-методичний комплекс
Форма проведення занять	Лекції, лабораторні роботи, самостійна робота
Семестровий контроль	Залік
Викладач	д.т.н., професор Новотарський М.А.