

**ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ І.І. МЕЧНИКОВА  
ФАКУЛЬТЕТ МАТЕМАТИКИ, ФІЗИКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ  
КАФЕДРА МАТЕМАТИЧНОГО ЗАБЕЗПЕЧЕННЯ КОМП'ЮТЕРНИХ СИСТЕМ**

**Силабус курсу « МЕТОДИ ОБРОБКИ ТЕКСТІВ ПРИРОДНОЇ МОВИ »**

|                           |   |
|---------------------------|---|
| <b>Обсяг</b>              | загальна кількість: кредитів – 3; годин – 90; змістовних модулів – 2  |
| <b>Семестр</b>            | 6 (2) весняний  |
| <b>Дні, Час, Місце</b>    | за розкладом занять   |
| <b>Викладач(и)</b>        | Пенко Валерій Георгійович, кандидат технічних наук, доцент кафедри математичного забезпечення комп'ютерних систем |
| <b>Контактний телефон</b> | (067)7441326  |
| <b>E-mail</b>             | <a href="mailto:vpenko@onu.edu.ua">vpenko@onu.edu.ua</a>  |
| <b>Робоче місце</b>       | кафедра математичного забезпечення комп'ютерних систем  |
| <b>Консультації</b>       | очні консультації: за розкладом занять<br>on-line консультації: ZOOM (посилання генерується на початку занять)    |

### **КОМУНІКАЦІЯ**

Комунікація зі студентами буде здійснюватися електронною поштою, в аудиторії або через ZOOM.

### **АНОТАЦІЯ КУРСУ**

Предметом вивчення курсу є основні типи завдань, що виникають в контексті обробки текстів природної мови, теоретичні методи та програмні засоби їх вирішення.

#### ***Пререквізити курсу***

Матеріал курсу ґрунтується на раніше отриманих студентами знаннях, практичних вміннях та навичках з тем та напрямів щодо алгоритмів, структур даних, дискретної математики та теорії ймовірностей. Відповідні курси викладаються у межах освітньої програми першого (бакалаврського) рівня вищої освіти за спеціальністю 123 «Комп'ютерна інженерія» перед вивчення курсу «Методи обробки текстів природної мови».

#### ***Постреквізити курсу***

За використаними теоретичними підходами та моделями цей курс доповнює дисципліну «Методи та алгоритми обробки зображень і комп'ютерний зір» в сфері аналізу та обробки даних і є базою для засвоєння наступних дисциплін освітньо-професійної програми підготовки бакалаврів за спеціальністю 123 «Комп'ютерна інженерія»: «Проектно-технологічна практика», «Переддипломна практика», «Виконання кваліфікаційної роботи бакалавра».

**Метою курсу** є вивчення основних сучасних підходів для вирішення базових завдань обробки текстів на природній мові та практичне застосування мови та бібліотек Python для вирішення цих завдань.

#### ***Зміст курсу***

На протязі вивчення курсу розглядаються наступні теми:

- Обчислення над мовою - проста статистика.
- Класифікація завдань обробки текстів природної мови.
- Отримання доступу до корпусів текстів і лексичних ресурсів.
- Використання основних лексичних ресурсів.
- Корпус WordNet.
- Доступ до тексту з Web і до локального тексту.
- Трубопровід NLP. Реалізація окремих етапів.

### **ОЧІКУВАНІ РЕЗУЛЬТАТИ**

У результаті вивчення курсу студент повинен

**знати:** основні можливості мови Python для розробки програмного забезпечення обробки текстів; основні можливості спеціалізованих пакетів для обробки текстів на природній мові; основні типи завдань, що стосуються обробки текстів на природній мові; особливості корпусо-орієнтованого підходу до обробки текстів природної мови.

**вміти:** розробляти програмне забезпечення, що виконує базові операції з текстами; використовувати спеціалізовані пакети Python для підвищення ефективності вирішення базових завдань обробки текстів на природній мові; застосовувати декілька різновидів мовних корпусів в якості ресурсу для вирішення завдань обробки текстів на природній мові.

**Компетентності**, які отримує студент у результаті вивчення курсу:

- Здатність до абстрактного мислення, аналізу і синтезу.
- Здатність вчитися і оволодівати сучасними знаннями.
- Вміння виявляти, ставити та вирішувати проблеми.
- Здатність застосовувати базові знання з фундаментальної та прикладної математики в професійній діяльності.
- Здатність використовувати сучасні методи і мови програмування для розроблення алгоритмічного та програмного забезпечення.
- Здатність використовувати та впроваджувати нові технології, включаючи технології розумних, мобільних, зелених і безпечних обчислень, брати участь в модернізації та реконструкції комп'ютерних систем та мереж, різноманітних вбудованих і розподілених додатків, зокрема з метою підвищення їх ефективності.
- Здатність вирішувати проблеми у галузі комп'ютерних та інформаційних технологій, визначати обмеження цих технологій.
- Здатність аргументувати вибір методів розв'язування спеціалізованих задач, критично оцінювати отримані результати, обґрунтовувати та захищати прийняті рішення.
- Здатність застосовувати закономірності випадкових явищ, ймовірно-статистичні методи, основи теорії чисельних методів та сучасні методи дискретної математики для аналізу і синтезу складних систем, методи кількісної оцінки інформації і створення коригуючих кодів при розв'язанні прикладних і наукових завдань в області комп'ютерної інженерії.
- Здатність використовувати декларативну парадигму програмування та мови, підходи, методи і технології штучного інтелекту, технології інженерії знань, інструментальні засоби підтримки інтелектуальних систем, розробляти та застосовувати моделі представлення знань, стратегії логічного виведення.

**Результати навчання:** по завершенню курсу студент матиме навички

- Мати навички проведення експериментів, збирання даних та моделювання в комп'ютерних системах.
- Знати новітні технології в галузі комп'ютерної інженерії.
- Вміти застосовувати закономірності випадкових явищ, ймовірно-статистичні методи, основи теорії чисельних методів та сучасні методи дискретної математики для аналізу і синтезу складних систем, методи кількісної оцінки інформації і створення коригуючих кодів при розв'язанні прикладних і наукових завдань в області комп'ютерної інженерії.
- Вміти застосовувати знання для ідентифікації, формулювання і розв'язування технічних задач спеціальності, використовуючи методи, що є найбільш придатними для досягнення поставлених цілей.
- Вміти розв'язувати задачі аналізу та синтезу засобів, характерних для спеціальності.
- Вміти здійснювати пошук інформації в різних джерелах для розв'язання задач комп'ютерної інженерії.
- Вміти виконувати експериментальні дослідження за професійною тематикою.

- *Вміти оцінювати отримані результати та аргументовано захищати прийняті рішення.*
- *Розробляти та застосовувати моделі представлення знань, стратегії логічного виведення, технологій інженерії знань, технологій і інструментальних засобів побудови інтелектуальних систем і систем штучного інтелекту.*

## **ФОРМИ І МЕТОДИ НАВЧАННЯ**

Курс буде викладений у формі лекцій (18 год.) та лабораторних занять (16 год.), організації самостійної роботи студентів (56 год.).

Основна підготовка студентів здійснюється на лекційних та лабораторних заняттях.

Під час викладання курсу використовуються такі **методи навчання**: *словесні* (лекція, пояснення); *наочні* (презентація Power Point); *практичні* (лабораторні роботи); *робота з літературними джерелами* (самостійна робота студентів).

### **Відповідність змісту курсу цілям сталого розвитку (з точки зору глобальних цілей сталого розвитку до 2030 року)**

.ЦСР 4: Якісна освіта - Викладання методів обробки текстів природної мови забезпечує студентів сучасними та практично важливими знаннями та навичками, необхідними для успішної кар'єри в галузі обробки даних та штучного інтелекту. Це сприяє підвищенню рівня освіти та професійної підготовки.

ЦСР 8: Гідна праця та економічне зростання - Випускники програм, що мьстять курс з обробки текстів природної мови, можуть знайти гідну роботу в високотехнологічних секторах, що сприяє економічному зростанню та зниженню рівня безробіття.

ЦСР 9: Індустріалізація, інновації та інфраструктура - Методи обробки текстів природної мови відіграють ключову роль у розвитку інновацій та технологій, які покращують інфраструктуру та сприяють індустріалізації.

ЦСР 17: Партнерство в інтересах сталого розвитку – Знання та навички при засвоєнні цього курсу сприяють міжнародному співробітництву та обміну знаннями, що важливо для досягнення всіх цілей сталого розвитку.