



Основи обробки природної мови
Робоча програма навчальної дисципліни (Силлабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	Перший (бакалаврський)
Галузь знань	12 «Інформаційні технології»
Спеціальність	122 «Комп'ютерні науки»
Освітня програма	«Системи та методи штучного інтелекту»
Статус дисципліни	вибіркова
Форма навчання	очна(денна)
Рік підготовки, семестр	4 курс, весняний семестр
Обсяг дисципліни	120 годин
Семестровий контроль/ контрольні заходи	залік
Розклад занять	Лекція раз на тиждень, практичні:раз на 2 тижні
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор, Практичні: к.т.н. Шаповал Наталія Віталіївна, shapoval.nataliia@lil.kpi.ua
Розміщення курсу	classroom

Програма навчальної дисципліни

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

Метою даної дисципліни є підготовка фахівців з комп'ютерних наук, здатних розв'язувати задачі з області обробки природної мови (ОПМ). Під час вивчення дисципліни формується вміння використовувати методології ОПМ для вирішення складних проблем у новому чи незнайомому середовищі, зокрема в соціальному, діловому та економічному контексті, аналізувати та обробляти великі дані за допомогою передових методів інтелектуального аналізу даних, специфічних для ОПМ, покращуючи процес прийняття стратегічних рішень.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Курс “Основи обробки природної мови” є вибіркоким курсом. Цей курс спирається на раніше прочитані спеціальні дисципліни в напрямку аналізу даних і дає систематизоване детальне викладання основ обробки природної мови та аналізу тексту. Ця дисципліна має логічні зв'язки з попередніми дисциплінами навчального плану підготовки, зокрема з курсами “Математичний аналіз”, “Математична статистика”, “Алгоритмізація та програмування”.

3. Зміст навчальної дисципліни

Розділ 1. Класифікація текстів

Тема 1.1. Структура тексту. Аналіз залежностей

Тема 1.2. Регулярні вирази. Аналіз настроїв

Тема 1.3. Токенізація, стемінг та лематизація

Тема 1.4. Статистичний підхід. N-грам моделі

Тема 1.5. Метрики оцінки моделей

Розділ 2. Представлення слів

Тема 2.1. Семантичні вектори та вбудування

Тема 2.2. Статистичне вбудування слів (Word2Vec, GloVe, FastText)

Тема 2.3. Контекстні вбудування (ELMo, BERT)

Тема 2.4. Визначення подібності текстів

Тема 2.5. Стилзація тексту

Розділ 3. Глибокі нейронні мережі в задачах обробки природної мови

Тема 3.1. Рекурентні нейронні мережі та трансформери

Тема 3.2. Архітектура BERT

Тема 3.3. Тонке налаштування моделей

Тема 3.4. Генерація тексту

Тема 3.5. Аналіз настроїв та класифікація текстів з використанням нейронних мереж

Розділ 4. Застосування обробки природної мови

Тема 4.1. Машинний переклад

Тема 4.2. Відповіді на запитання та пошук інформації

Тема 4.3. Чат-боти та системи діалогу

Тема 4.4. Автоматичне розпізнавання мовлення та перетворення тексту в мовлення

Розділ 5. Великі мовні моделі

Тема 5.1. Великі мовні моделі. Вступ

Тема 5.2. Метрики оцінки моделей

Тема 5.3. Проблема галюцинацій

4. Навчальні матеріали та ресурси

Основна:

1. [JURAFSKY, Daniel; MARTIN, James H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.*](#)
2. [GOLDBERG, Yoav. *Neural network methods for natural language processing.* Springer Nature, 2022.](#)

Додаткова

1. Статті які будуть додані в класрумі

Навчальний контент

5. Методика опанування навчальної дисципліни(освітнього компонента)

Лекційні заняття містять детальний огляд тем з Розділів 1-5, а також практичні вправи для розв'язку задач, що розглядаються на лекціях. Передбачається попередній розгляд теми та вправ самостійно перед лекцією для кращого засвоєння матеріалу.

Лабораторні роботи

№ п/п	Найменування лабораторної роботи	Кількість годин
ЛР №1	Розпізнавання іменованих сутностей	2
ЛР №2	Класифікація текстів	4
ЛР №3	Машинний переклад	4
ЛР №4	Створення чат-боту	4
ЛР №5	Дотренування великих мовних моделей	4

6. Самостійна робота студента

Самостійна робота включає такі активності, як підготовка до лекційних та лабораторних аудиторних занять, ознайомлення з порядком виконання та змістом лабораторних робіт, формування звітів, підготовка до захисту лабораторних робіт, виконання додаткових завдань для кращого опанування матеріалу, підготовка до написання контрольної роботи з дисципліни.

Контрольна робота. Робочим навчальним планом передбачено виконання контрольної роботи, яка ділиться на 2 частини.

Політика та контроль

6. Політика навчальної дисципліни (освітнього компонента)

Здобувачі повинні відвідувати всі заняття, які проводяться за розкладом (у контактній або дистанційній формах). Викладач контролює присутність студентів на заняттях. Відвідування консультаційних занять з дисципліни не є обов'язковим.

Під час проведення лекційних занять студенту рекомендується вести докладний конспект лекцій, приймати активну участь у обговоренні проблемних ситуацій, відповідати на питання, поставлені викладачем.

Перед проведенням лабораторних робіт студенти повинні підготувати (засвоїти) теоретичний лекційний матеріал за відповідною темою, уявити порядок виконання лабораторної роботи, визначити або згенерувати (якщо це передбачено) індивідуальний варіант (варіанти) завдання. При виконанні лабораторної роботи дозволяється використовувати стандартні пакети прикладних програм або математичні середовища. При цьому неприпустимим є використання таких пакетів або середовищ, які захищені правом інтелектуальної власності.

Також студенти можуть писати власний програмний код. Після виконання завдань, передбачених в лабораторній роботі, студенти повинні провести аналіз отриманих результатів, сформулювати звіт у відповідності до висунутих вимог, підготуватися до захисту лабораторної роботи.

Лабораторні роботи виконуються індивідуально. Якість виконання і захисту лабораторних робіт оцінюється згідно рейтингової системи. Для виконання і захисту кожної лабораторної роботи встановлюються дедлайни.

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

Поточний контроль проводиться при виконанні і захисті лабораторних робіт. Оцінюється теоретична підготовка (обізнаність щодо застосування методів обробки природної мови), а також володіння практичними навичками (реалізація та аналіз розв'язку поставленої задачі). За несвоєчасну здачу лабораторних робіт нараховуються штрафні бали. Система рейтингових (вагових) балів при виконанні поточного контролю наведена у табл. 7.1.

Також є можливість отримувати додаткові бали за необов'язкові додаткові завдання та виконання індивідуальної творчої роботи.

Календарний контроль проводиться двічі на семестр як моніторинг поточного стану виконання вимог силлабусу.

Семестровий контроль – залік. Умови допуску до семестрового контролю: зарахування усіх лабораторних робіт та загальний семестровий рейтинг не менше 35 балів. Відповідність рейтингових балів оцінкам за університетською шкалою наведено у табл. 7.2.

Таблиця 7.1. Система рейтингових (вагових) балів та критеріїв оцінювання

Категорія оцінювання	Мінімальна оцінка в балах	Максимальна оцінка в балах
Лабораторна робота 1	7	11
Лабораторна робота 2	7	11
Лабораторна робота 3	7	11
Лабораторна робота 4	7	11
Лабораторна робота 5	7	11
Тести	0	15
Контрольна робота 1	0	15
Контрольна робота 2	0	15
Разом	35	100

Таблиця 7.2. Відповідність рейтингових балів оцінкам за університетською шкалою

Загальна сума балів	Оцінка
100...95	Відмінно
94...85	Дуже добре
84...75	Добре
74...65	Задовільно
64...60	Достатньо
Менше 60	Незадовільно

стартовий рейтинг менше 30 балів	Не допущено
-------------------------------------	-------------

Критерії нарахування балів за контрольні заходи:

~ “відмінно”: 95 -100% - здобувач виявив всебічні, систематичні та глибокі знання навчального матеріалу з дисципліни; продемонстрував уміння вільно виконувати всі завдання, передбачені програмою; засвоїв основну та додаткову літературу; проявив творчі здібності в розумінні, в логічному, чіткому, стислому та ясному трактуванні навчального матеріалу; засвоїв взаємозв’язок основних понять дисципліни, їх значення для подальшої професійної діяльності

~ “дуже добре”: 85-94% - здобувач виявив систематичні знання навчального матеріалу з дисципліни вище середнього рівня; продемонстрував уміння добре виконувати всі завдання, передбачені програмою, допустивши незначні помилки; засвоїв основну та додаткову літературу; засвоїв взаємозв’язок основних понять дисципліни, їх значення для подальшої професійної діяльності

~ “добре”: 75-84% - здобувач виявив загалом добрі знання навчального матеріалу при виконанні передбачених програмою завдань, але припустив ряд помітних помилок; засвоїв основну літературу; показав систематичний характер знань з дисципліни; здатний до їх самостійного використання та поповнення в процесі подальшої навчальної роботи і професійної діяльності

~ “задовільно”: 65-74% - здобувач виявив знання основного навчального матеріалу з дисципліни в обсязі, необхідному для подальшого навчання та майбутньої професійної діяльності; ознайомився з основною літературою; впорався з виконанням завдань, передбачених програмою, але припустив значну кількість помилок або недоліків на запитання при співбесіді, тестуванні та при виконанні завдань тощо, принципів з яких може усунути самостійно

~ “достатньо”: 60-64% - здобувач виявив знання основного навчального матеріалу з дисципліни в мінімальному обсязі, необхідному для подальшого навчання та майбутньої професійної діяльності; ; ознайомився з основною літературою; в основному виконав завдання, передбачені програмою, але припустив помилки у відповіді на запитання при співбесідах, тестуванні та при виконанні завдань тощо, які він може усунути лише під керівництвом та за допомогою викладача

~ “незадовільно”: 0-59% - здобувач мав значні прогалини в знаннях основного навчального матеріалу; допускав принципові помилки при виконанні передбачених програмою завдань, але спроможній самостійно доопрацювати програмний матеріал і підготуватися для перездачі дисципліни

8. Додаткова інформація з дисципліни (освітнього компонента)

Надається можливість зарахування сертифікатів проходження дистанційних чи онлайн курсів за відповідною тематикою у вигляді додаткових балів чи зарахуванні певної теми.

Робочу програму навчальної дисципліни (силлабус):

Складено доцент, к.т.н., Шаповал Наталія Віталіївна

Ухвалено кафедрою штучного інтелекту (протокол № 14 від 11.06.2024)

Погоджено Методичною комісією ННІПСА (протокол № 10 від 24.06.2024)