



Методи обробки природної мови

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>121 Інженерія програмного забезпечення 123 Комп'ютерна інженерія 126 Інформаційні системи та технології</i>
Освітня програма	<i>Інженерія програмного забезпечення комп'ютерних систем Інженерія програмного забезпечення інформаційних систем Комп'ютерні системи та мережі Інтегровані інформаційні системи Інформаційні управляючі системи та технології Інформаційне забезпечення робототехнічних систем Інформаційні системи та технології</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>Очна(денна)</i>
Рік підготовки, семестр	<i>2 курс, 3-й семестр</i>
Обсяг дисципліни	<i>4 кредити/120 год. денна форма: лекцій 36 год., лабораторних занять 18 год., СРС 66 год.</i>
Семестровий контроль/ контрольні заходи	<i>Залік</i>
Розклад занять	<i>schedule.kpi.ua</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу/ викладачів	<i>Лектор: ст. викладач Кочура Юрій Петрович, iuriy.kochura@gmail.com Лабораторні роботи: ст. викладач Кочура Юрій Петрович, iuriy.kochura@gmail.com</i>
Розміщення курсу	<i>https://classroom.google.com/c/ODM2ODcyNDkwODgz?cjc=afbp7yl</i>

Програма навчальної дисципліни

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

Дисципліна присвячена вивченню методів аналізу, моделювання та генерації текстових даних. Студенти познайомляться з сучасними підходами, які використовуються для представлення слів і речень, моделювання мови, нейронними мережами для обробки природної мови, механізмом уваги та трансформерами.

Предмет навчальної дисципліни – алгоритми і моделі обробки природної мови.

Метою дисципліни є формування у студентів системних знань і практичних умінь щодо розробки та застосування алгоритмів обробки природної мови у наукових і прикладних задачах.

Вивчення дисципліни формує наступні **компетенції**:

-Здатність застосовувати сучасні моделі обробки природної мови для аналізу, класифікації та генерації тексту.

-Здатність організовувати ефективну обробку текстових даних.

Програмні результати навчання, на формування яких спрямована дисципліна:

1. Уміти застосовувати векторні представлення слів і речень, нейронні моделі мови та трансформери для аналізу, класифікації та генерації тексту.
2. Уміти аналізувати продуктивність та ефективність моделей NLP.
3. Уміти використовувати методи перенесення знань для розв'язання практичних задач обробки природної мови.

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

Пререквізити

Для опанування дисципліни необхідні знання з лінійної алгебри та вищої математики, уміння програмувати на Python, а також розуміння основ алгоритмів і структур даних.

Постреквізити

Отримані знання та практичні навички будуть корисні для розв'язання прикладних задач обробки тексту, машинного перекладу, класифікації та генерації тексту, а також для роботи з моделями-основами і трансформерами.

Знання, одержані студентами під час вивчення дисципліни можуть бути використані у дисциплінах «Науково-дослідна практика», «Машинне навчання», «Нейронні мережі» та інші.

3. Зміст навчальної дисципліни

Розділ 1. Векторні представлення слів

Тема 1.1. Векторні подання слів та розподілена семантика

Тема 1.2. Класичні методи підрахунку частот

Тема 1.3. Моделі Word2Vec та GloVe

Тема 1.4. Оцінка векторних моделей та їх інтерпретація

Розділ 2. Моделювання мови

Тема 2.1. Поняття моделювання мови

Тема 2.2. Моделі на основі N-gram

Тема 2.3. Нейронні моделі мови: рекурентні та згорткові

Розділ 3. Seq2seq та механізм уваги

Тема 3.1. Основи Seq2seq: архітектура “кодер-декодер”

Тема 3.2. Механізм уваги: загальний опис та різні функції оцінки

Тема 3.4. Архітектура трансформера: самоувага, багатоголовий механізм

Тема 3.5. Токенізація на рівні субслів (BPE-Byte-Pair Encoding)

Розділ 4. Перенесення знань

Тема 4.1. Основи та принципи перенесення знань у NLP

Тема 4.2. Контекстні векторні подання слів (CoVe, ELMo)

Тема 4.3. Попередньо навчені трансформери та адаптація моделей (GPT, BERT)

Розділ 5. Великі мовні моделі

Тема 5.1. Закони масштабування та виникнення нових можливостей

Тема 5.2. Техніки роботи з запитами (“prompting”)

Тема 5.3. Гіпотези щодо роботи моделей при навчанні на контексті

4. Навчальні ресурси та матеріали

Базова література

1. Методи обробки природної мови: Феномен ChatGPT [Текст]: навч. посіб. для здобувачів ступеня бакалавра за освітньою програмою «Інженерія програмного забезпечення інтелектуальних кібер-фізичних систем в енергетиці» / автор:Ю.І.Стативка; КПІ ім. Ігоря Сікорського.– Електронні текстові дані (1 файл: 621 кбайт). – Київ: КПІ ім. Ігоря Сікорського, 2023. – 67 с.

2. D. Jurafsky, J.H. Martin: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 3rd ed. https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf

Додаткова література

3. K. Bhavsar, N. Kumar, P. Dangeti: Natural Language Processing with Python Cookbook. // Packt Publishing, 2017.

4. Lewis Tunstall. Natural Language Processing with Transformers, Revised Edition / Lewis Tunstall, Leandro von Werra, Thomas Wolf. – Beijing-Boston-Farnham-Sebastopol-Tokyo: O'Reilly Media, Inc., 2022. – 383 p.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компоненту)

Лекційні заняття

№ розділу	Назва лекції та перелік основних питань	Кількість ауд. годин
1	Розділ 1. Векторні представлення слів Тема 1.1. Векторні подання слів та розподілена семантика Тема 1.2. Класичні методи підрахунку частот Тема 1.3. Моделі Word2Vec та GloVe Тема 1.4. Оцінка векторних моделей та їх інтерпретація	8
2	Розділ 2. Моделювання мови Тема 2.1.Поняття моделювання мови Тема 2.2. Моделі на основі N-gram Тема 2.3. Нейронні моделі мови: рекурентні та згорткові	6
3	Розділ 3. Seq2seq та механізм уваги Тема 3.1.Основи Seq2seq: архітектура “кодер-декодер” Тема 3.2. Механізм уваги: загальний опис та різні функції оцінки Тема 3.4. Архітектура трансформера: самоувага, багатоголовий механізм Тема 3.5.Токенізація на рівні субслів (BPE-Byte-Pair Encoding)	8
4	Розділ 4. Перенесення знань Тема 4.1. Основи та принципи перенесення знань у NLP	

	Тема 4.2. Контекстні векторні подання слів (CoVe, ELMo) Тема 4.3. Попередньо навчені трансформери та адаптація моделей (GPT, BERT)	8
5	Розділ 5. Великі мовні моделі Тема 5.1. Закони масштабування та виникнення нових можливостей Тема 5.2. Техніки роботи з запитами (“prompting”) Тема 5.3. Гіпотези щодо роботи моделей при навчанні на контексті	6
	Разом:	36

Лабораторні заняття

№ з/п	Назва лабораторної роботи	Кількість ауд. годин
1	Векторне подання слів	4
2	Мовні моделі та N-gram	4
3	Seq2Seq моделі для перекладу	6
4	Тонке налаштування BERT для задачі класифікації тексту	4
	Разом:	18

Модульна контрольна робота

Модульна контрольна робота виконується за усіма темами, що розглядаються в дисципліні.

6. Самостійна робота здобувача вищої освіти денної форми навчання

До СРС відносяться такі види робіт:

- вивчення теоретичного матеріалу лекційних занять;
- виконання завдань лабораторних робіт та оформлення протоколів;
- підготовка до модульної контрольної роботи та семестрового контролю.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Академічна доброчесність

Під час проходження цього курсу студенти зобов'язані дотримуватись політики та принципів академічної доброчесності визначених Кодексом честі КПП ім. Ігоря Сікорського (Детальніше Розділ 3: <https://kpi.ua/code>) та усі наступні правила:

1. Кожен із здобувачів повинен відправляти на перевірку власно виконану роботу. Використання чужих розв'язків або програмного коду і представлення їх за свої напрацювання є плагіатом та

серйозним порушенням основних академічних стандартів.

2. Здобувачі не повинні ділитися своїми розв'язками з іншими студентами, а також просити інших ділитися своїми розв'язками з Вами.

3. Якщо студенти отримували допомогу у вирішенні певного завдання — потрібно зазначити це у звіті, а саме: від кого та яку допомогу отримували.

Норми етичної поведінки

Норми етичної поведінки студентів і працівників університету визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Виконання лабораторних робіт

Викладачем встановлюється граничний термін для виконання кожного практичного завдання з метою послідовного та збалансованого засвоєння студентами навчального матеріалу, визначеного даним силабусом. Здача на перевірку практичних завдань після встановлених термінів супроводжується штрафними балами. Кінцевим терміном для здачі на перевірку усіх практичних завдань є останнє заняття з комп'ютерного практикуму. У період проведення в університеті семестрового контролю приймання практичних завдань не здійснюється. Студент(-ка) має можливість переробити або доопрацювати практичне завдання на перездачі з метою підвищення свого поточного рейтингу.

Процедура оскарження оцінок

Якщо студент(-ка) вважає, що його(її) роботу недооцінено або переоцінено — потрібно звернутися до викладача, який здійснював оцінювання та повідомити про це із зазначенням короткого обґрунтування (оцінка буде переглянута).

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

Поточний контроль

Поточний контроль успішності засвоєння знань виконується шляхом перевірки завдань лабораторних робіт та МКР. Таким чином, семестровий рейтинг студентів з дисципліни складається з балів, що вони отримують за:

- виконання захист завдань лабораторних робіт;
- модульну контрольну роботу.

Календарний контроль

Календарний контроль провадиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

Умови позитивного календарного контролю:

- за результатами навчальної роботи на першому календарному контролі (7-й тиждень семестру) студент отримує статус «атестований», якщо його поточний рейтинг становить не менше ніж 50% від максимально можливої кількості балів, які можна було набрати протягом перших 6 тижнів;
- за результатами навчальної роботи на другому календарному контролі (13-й тиждень семестру) студент отримує статус «атестований», якщо його поточний рейтинг становить не менше ніж 50% від максимально можливої кількості балів, які можна було набрати протягом перших 12 тижнів.

Семестровий контроль

Семестровий контроль проводиться для встановлення рівня досягнення здобувачами програмних результатів навчання з навчальної дисципліни (освітнього компонента) за семестр. Семестровий контроль проводиться відповідно до навчального плану у вигляді заліку в терміни, встановлені графіком навчального процесу.

Здобувач допускається до семестрового контролю, якщо:

- виконаві захистив усі передбачені силабусом лабораторні роботи;
- набрав не менше ніж **60%** від максимальної кількості балів поточного контролю:

Важливо! Умова допуску до семестрового контролю (заліку):
Лабораторні роботи+МКР $\geq$ 36 балів

Система рейтингових (вагових) балів та критерії оцінювання

Система оцінювання

40 балів	Лабораторні роботи (10 балів кожна)
20 балів	МКР
40 балів	Залік

Завдання лабораторних робіт

Ваговий бал кожної лабораторної роботи—10 балів. Максимальна кількість балів за всі лабораторні роботи дорівнює: $4 * 10 = 40$ балів.

Складові оцінки та відповідний максимальний бал:

- програмна реалізація алгоритмів—2 бали;
- експериментальне дослідження / порівняння алгоритмів—2 бали;
- звіт (відповідність структурі, презентація результатів роботи, якість та обґрунтування рішень, висновки) – 2 бали;
- захист результатів комп'ютерного практикуму (теоретичні знання за темою лабораторної роботи; обґрунтування прийнятих рішень; правильні відповіді на питання) – 4 бали.

Бал може бути знижений у разі неповного виконання завдань, недостатньої якості звіту, помилоку програмному коді чи логіці реалізації, неправильної інтерпретації отриманих результатів, порушення принципів академічної доброчесності або строку здачі завдань.

Модульна контрольна робота (МКР)

Модульна контрольна робота проводиться у вигляді тесту з 10 питань. Ваговий бал кожного питання становить 2 бали. Правильна відповідь – 2 бали; неправильна – 0 балів. Максимальна кількість балів за МКР – 20 балів.

Залікова робота

Максимальна кількість балів, які можна отримати за результатами здачі заліку—40 балів.

Складові оцінки та відповідний максимальний бал за виконання залікової роботи:

- теоретичні питання (розуміння ключових понять, методів і принципів, здатність пояснити теоретичні положення власними словами, правильність використання термінології) –15 балів;
- здатність застосувати теоретичні знання до практичних завдань (коректність побудови алгоритмів для розв'язку задач, правильність обчислень, обґрунтованість отриманих результатів) – 20 балів;
- додаткові уточнювальні запитання (розуміння отриманих результатів, уміння аргументувати обраний підхід, здатність аналізувати помилки та пояснювати альтернативні рішення) – 5 балів.

Максимальна сума вагових балів контрольних заходів протягом семестру складає:

$$4 \cdot 10 + 20 + 40 = 100.$$

Сума балів, які отримують студенти за роботу протягом семестру, обчислюється за формулою:

$$R_D = \sum_{i=1}^4 r_{ЛРi} + r_{МКР} + r_{ЗЛ}$$

де $r_{ЛРi}$ – бали, отримані за i -ту лабораторну роботу (0...10);

$r_{МКР}$ – бали, отримані за МКР (0...20).

$r_{ЗЛ}$ – бали, отримані на заліку (0...40).

Переведення рейтингових балів до оцінок за університетською шкалою (<https://kpi.ua/grading>):

95–100	Відмінно
85–94	Дуже добре
75–84	Добре
65–74	Задовільно
60–64	Достатньо
Менше 60	Незадовільно
Менше 36	Не допущений
Порушення принципів академічної доброчесності або морально-етичних Норм поведінки	Усунений

9. Нормативні положення політики навчальної дисципліни (освітнього компонента)

Правила відвідування занять. Відвідування занять є вільним, бали за присутність на лекціях та практичних заняттях не нараховуються, відповідно штрафні бали у разі відсутності на занятті не передбачаються.

Дистанційний режим навчання. У разі запровадження обмежень на відвідування університету, пов'язаних з введенням карантину або режиму воєнного стану в державі, освітній процес здійснюється у дистанційному режимі відповідно до Положення про дистанційне навчання в КПІ ім. Ігоря Сікорського (<https://osvita.kpi.ua/index.php/node/188>), Регламенту організації освітнього процесу в дистанційному режимі (<https://proffkom.kpi.ua/reglament-organizatsiyi-osvitnogo-protsesu-v-distantsiynomu-rezhimi>) та Регламенту проведення семестрового контролю в дистанційному режимі (<https://osvita.kpi.ua/node/148>). У режимі дистанційного навчання заняття відбуваються у вигляді онлайн-конференції на платформах BigBlueButton, Google Meet, 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://www.sikorsky-distance.org>). Результати оцінювання висвітлюють у АС «Електронний кампус» на особистій сторінці здобувача (<https://ecampus.kpi.ua>).

Правила поведінки на заняттях. На заняттях слід дотримуватись норм етичної поведінки визначених у Кодексі честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>). На території університету здобувачі мають поводити себе відповідно до Правил внутрішнього розпорядку (<https://kpi.ua/admin-rule>).

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 навчання на неформальній освіті, яку студент обрав самостійно і хоче зарахувати результати навчання в рамках дисципліни, 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.

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, такі індивідуальні завдання, мають бути результатом власної оригінальної роботи здобувача. Використання ШІ для автоматичної генерації відповідей без подальшого їх аналізу та доопрацювання заборонено.

Робочу програму навчальної дисципліни (силабус)

Складено, ст.викладач кафедри ОТ Кочура Ю.П.,

Ухвалено: кафедрою обчислювальної техніки (протокол №12 від 23.06.2025р.)

Погоджено: методичною комісією ФІОТ (протокол №11 від 27.06.2025р.)