

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
Кафедра оптоелектроніки та інформаційних технологій

Затверджено

На засіданні
кафедри оптоелектроніки та
інформаційних технологій
факультету електроніки та комп'ютерних
технологій
Львівського національного університету
імені Івана Франка
(протокол №6 від 29.08. 2023 р.)

Завідувач кафедри:



Олег КУШНІР

Силабус з навчальної дисципліни
“Методи та технології CV та NLP”,
що викладається в межах ОПП
“ Високопродуктивний комп'ютинг ”
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 121 – Інженерія програмного забезпечення

Львів 2023 р.

Назва дисципліни	Методи та технології CV та NLP
Адреса викладання дисципліни	Корпус факультету електроніки та комп'ютерних технологій, Львівський національний університет імені Івана Франка, вул. Драгоманова 50, м. Львів, 79005, вул. Ген. Тарнавського 107, м. Львів, 79011
Факультет та кафедра, за якою закріплена дисципліна	Факультет електроніки та комп'ютерних технологій Кафедра оптоелектроніки та інформаційних технологій
Галузь знань, шифр та назва спеціальності	12 –Інформаційні технології 121–Інженерія програмного забезпечення
Викладачі дисципліни	Кушнір О.С., докт. фіз.-мат. наук, професор, професор Дуфанець М.В., доктор філософії (прикладна фізика та наноматеріали)
Контактна інформація викладачів	oleh.kushnir@lnu.edu.ua https://electronics.lnu.edu.ua/employee/kushnir-o-s marta.dufanets@lnu.edu.ua , https://electronics.lnu.edu.ua/employee/dufanets-marta-vasylyvna
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекційних/лабораторних занять (за попередньою домовленістю): ауд. 211, корпус факультету електроніки та комп'ютерних технологій, вул. Тарнавського 107, м. Львів
Сторінка дисципліни	https://e-learning.lnu.edu.ua/course/view.php?id=6015
Інформація про дисципліну	Дисципліна “Методи та технології CV та NLP” є вибірковою дисципліною зі спеціальності 121–Інженерія програмного забезпечення для освітньої програми “Високопродуктивний комп'ютинг”, яка викладається у 8 семестрі в обсязі 5 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Навчальну дисципліну розроблено для одержання здобувачами теоретичних знань з обробки природної мови, розпізнавання образів, методів визначення ключових ознак, їх аналізу, та віднесення їх до певного класу. Курс охоплює основні алгоритми комп'ютерного зору: попередню обробку, фільтрацію, сегментацію, виділення ознак, розпізнавання, класифікацію зображень, а також програмну реалізацію алгоритмів розпізнавання в різноманітних системах штучного інтелекту, розробку ефективного програмного забезпечення для вирішення стандартних прикладних задач опрацювання природної мови.
Мета та цілі дисципліни	<i>Метою</i> вивчення дисципліни “Методи та технології CV та NLP” є ознайомлення здобувачів з теоретичними основами розв'язання задачі розпізнавання образів, алгоритмами та методами опрацювання природної мови. <i>Цілями</i> є формування в них практичних навичок, які б дали змогу ефективно застосовувати засвоєні знання, алгоритми, методи та наявні бібліотеки та он лайн ресурси для вирішення таких задач.
Література для вивчення дисципліни	Основна література: 1. M. Schlesinger, V. Hlavac Ten Lectures on Statistical and Structural Pattern Recognition // Computational Imaging and Vision, Vol. 24. Kluwer Academic Publishers– Dordrecht / Boston / London. – 2002. – 520 p.

	<ol style="list-style-type: none"> 2. Reinhard Klette. Concise Computer Vision: An Introduction into Theory and Algorithms (Undergraduate Topics in Computer Science). - Springer - January 20th, 2014 - 429 p. 3. Муравський Л.І., Бобицький Я.В., Гаськевич Г.І. Оптичні інформаційні системи: Підручник. – Львів: СПОЛЮМ, 2011. – 200 с. 4. Русин Б.П. Структурно-лінгвістичні методи розпізнавання зображень в реальному часі. Київ, Наукова думка, 1986. - 128 с. 5. Волошин В. Г. Комп'ютерна лінгвістика / В. Г. Волошин. – Суми : Університетська книга, 2004. – 382 с. 6. Кушнір О. С. Основи комп'ютерної лінгвістики (конспект лекцій) / О. С. Кушнір. – Львів : Видавн. Львів. ун-ту, 2023. – 292 с. 7. Пасічник В. В. Математична лінгвістика. Книга 1. Квантитативна лінгвістика / В. В. Пасічник, Ю. М. Щербина, В. А. Висоцька, Т. В. Шестакевич. – Львів : Новий світ – 2000, 2012. – 359 с. 8. Clark A. The Handbook of Computational Linguistics and Natural Lan-guage Processing / A. Clark, C. Fox, S. Lappin. – Chichester : John Wiley and Sons, 2010. – 801 p. <p style="text-align: center;">Додаткова література:</p> <ol style="list-style-type: none"> 1. Reinhard Klette. Concise Computer Vision: An Introduction into Theory and Algorithms (Undergraduate Topics in Computer Science). - Springer - January 20th, 2014 - 429 p. 2. You can master Computer Vision, Deep Learning, and OpenCV. - Режим доступу: https://www.pyimagesearch.com 3. Manning C. D. Foundations of statistical natural language processing / Manning C. D., Schutze H. – London : The MIT Press Cambridge, 1999. – 680 p.
Обсяг курсу	Загальний обсяг: 150 годин. Аудиторних занять: 64 год., з них 32 год. лекційних та 32 годин лабораторних робіт. Самостійної роботи: 86 год.
Очікувані результати навчання	<p>У результаті вивчення даного курсу студент буде:</p> <p><i>знати:</i> основні принципи теорії розпізнавання образів за детерміністичного та імовірнісного підходів, особливості стохастичної класифікації та розпізнавання оптичних образів; основні теорії, моделі та алгоритми комп'ютерної лінгвістики і опису лінгвістичних систем, інформаційного пошуку та інтелектуального аналізу текстових даних;</p> <p><i>вміти</i> розв'язувати задачі класифікації образів з використанням розв'язувальних функцій та функцій відстані, використовуючи для цього відповідне програмне забезпечення, створювати програмні модулі на основі стандартних методів розпізнавання з бібліотеки OpenCV та користуватися ними; аналізувати моделі комп'ютерної лінгвістики, працювати з відповідними програмними продуктами, застосовувати комп'ютерну техніку для вирішення лінгвістичних задач.</p>

Після вивчення даного курсу “Методи та технології CV та NLP” здобувачі набудуть таких Загальних(ЗК)/Фахових(ФК) компетентностей та Програмних результатів навчання (ПРН):

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК03. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК04. Здатність спілкуватися іноземною мовою як усно, так і письмово.

ЗК06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ФК13. Здатність ідентифікувати, класифікувати та формулювати вимоги до програмного забезпечення.

ФК14. Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування.

ФК15. Здатність розробляти архітектури, модулі та компоненти програмних систем.

ФК16. Здатність формулювати та забезпечувати вимоги щодо якості програмного забезпечення у відповідності з вимогами замовника, технічним завданням та стандартами.

ФК19. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.

ФК21. Здатність оцінювати і враховувати економічні, соціальні, технологічні та екологічні чинники, що впливають на сферу професійної діяльності.

ФК24. Здатність здійснювати процес інтеграції системи, застосовувати стандарти і процедури управління змінами для підтримки цілісності, загальної функціональності і надійності програмного забезпечення.

ФК25. Здатність обґрунтовано обирати та освоювати інструментарій з розробки та супроводження програмного забезпечення.

ФК26. Здатність до алгоритмічного та логічного мислення.

ФК27. Здатність розробляти високопродуктивні програмні комплекси для вирішення задач наук про дані, систем штучного інтелекту, вбудованих та інших інноваційних систем.

ФК28. Володіння методами розроблення і впровадження систем підвищеної продуктивності, серверних, мікросервісних, хмаркових, розподілених та інших новітніх технологій.

ПРН03. Знати основні процеси, фази та ітерації життєвого циклу програмного забезпечення.

ПРН05. Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об’єктно-орієнтованого аналізу та математичного моделювання для розробки програмного забезпечення.

ПРН06. Вміння вибирати та використовувати відповідну задачі методологію створення програмного забезпечення.

ПРН07. Знати і застосовувати на практиці фундаментальні концепції, парадигми і основні принципи функціонування мовних, інструментальних і обчислювальних засобів інженерії програмного забезпечення.

	<p>ПРН09. Знати та вміти використовувати методи та засоби збору, формулювання та аналізу вимог до програмного забезпечення.</p> <p>ПРН10. Проводити передпроектне обстеження предметної області, системний аналіз об'єкта проектування.</p> <p>ПРН11. Вибирати вихідні дані для проектування, керуючись формальними методами опису вимог та моделювання.</p> <p>ПРН12. Застосовувати на практиці ефективні підходи щодо проектування програмного забезпечення.</p> <p>ПРН13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.</p> <p>ПРН20. Знати підходи щодо оцінки та забезпечення якості програмного забезпечення.</p> <p>ПРН25. Вміти застосовувати інноваційні технологічні рішення при розробці високопродуктивних систем.</p> <p>ПРН26. Знати засоби інтеграції, розгортання та підтримки спеціалізованих програмних компонентів, розроблених на основі інноваційних технологій для вирішення завдань високопродуктивних технологій.</p> <p>ПРН27. Знати основи інженерії й аналізу даних та вміти вибрати оптимальні алгоритми і технології для розробки інноваційних рішень при розв'язанні задач наук про дані, вбудованих систем та систем штучного інтелекту.</p> <p>ПРН28. Зберігати та примножувати цивілізаційні цінності і досягнення суспільства, діяти соціально відповідально та свідомо, зберігати навколишнє середовище, знати правила ведення здорового способу життя і надання першої медичної допомоги.</p>
Ключові слова	Розпізнавання образів, аналіз та обробка зображень, класифікація образів, статистична лінгвістика, опрацювання природної мови, машинний переклад, аналіз і синтез мови
Формат курсу	Очний
	Проведення лекцій, лабораторних робіт та консультації для кращого розуміння тем
Теми	Див. схему курсу
Підсумковий контроль, форма	Залік вкінці семестру
Пререквізити	Для вивчення курсу студенти потребують базових знань з дисциплін «Вища математика», «Дискретна математика», «Методи та технології обчислень», «Бази даних», «Об'єктно-орієнтоване програмування», «Прикладна статистика та ймовірнісні процеси».
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Інформаційні методи (лекція, бесіда, ілюстрація, демонстрація), дедуктивні методи на основі узагальнень, евристичні методи (проблемна лекція), інтерактивні методи, лабораторні роботи.
Необхідне обладнання	<p>Для проведення лекційних занять: комп'ютер (мінімальні характеристики: процесор Intel Core i3(4 ядра/8 потоків), 8ГБ оперативної пам'яті, 50ГБ вільного місця на диску, відеокарта Nvidia GEFORCE GT1030 2048MB), доступ до мережі Internet, засоби мультимедіа (в т.ч. проектор).</p> <p>Для проведення лабораторних занять: Комп'ютер (мінімальні характеристики: процесор Intel Core i3(4 ядра/8 потоків), 8ГБ</p>

	<p>оперативної пам'яті, 50ГБ вільного місця на диску, відеокарта Nvidia GEFORCE GT1030 2048MB). ОС Win10. MS Visual Studio 2017. Платформа Moodle та MS Teams.</p>
<p>Критерії оцінювання (окремо для кожного виду навчальної діяльності)</p>	<p>Оцінювання проводиться упродовж семестру за 100-бальною шкалою. Бали нараховуються за такими видами робіт з наступним співвідношенням:</p> <ul style="list-style-type: none"> • Змістовий модуль 1: до 25 балів за виконання модульного завдання. • Змістовий модуль 2: до 25 балів за виконання модульного завдання • Виконання лабораторних робіт: до 50 балів. <p>Підсумкова максимальна кількість балів 100. Звітність за курс – залік в кінці семестру.</p> <p>Академічна доброчесність: Очікується, що роботи студентів будуть оригінальними дослідженнями чи міркуваннями. Списування та втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в написанні завдань є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману. Жодні форми порушення академічної доброчесності не толеруються.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції і лабораторні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися усіх термінів визначених для виконання всіх видів робіт, передбачених курсом.</p> <p>Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p>Політика виставлення балів. Враховуються бали, набрані при поточному контролі та бали за виконання лабораторних робіт. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізнь на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін.</p> <p>Оцінювання лабораторних робіт (10 лабораторних робіт, максимальна кількість балів: 50) відбувається шляхом оцінки роботи студента під час проведення лабораторної роботи в аудиторії (0-5 балів за одну роботу) та захисту звіту по виконаній лабораторній роботі (0-5 балів за одну роботу). У підсумку, всі набрані бали множаться на коефіцієнт (0.5) для переведення у 50-ти бальну шкалу.</p> <p>Бали оцінювання лабораторних робіт нараховуються за наступним співвідношенням:</p> <p>5 – студент в повному обсязі володіє навчальним матеріалом, має повне розуміння розглянутої теми, надає правильні відповіді на запитання по темі, завдання виконане відповідно до вимог;</p>

	<p>4 – студент достатньо розуміє розглянутий матеріал та принципи виконаного завдання, присутні неточності та незначні помилки у відповідях на запитання по темі;</p> <p>3 – студент не досить добре розуміє розглянутий матеріал та зміст виконаного завдання, вагається та надає неточні/не конкретні відповіді на запитання по темі;</p> <p>2 – студент погано розуміє розглянутий матеріал, студент в більшості надає помилкові відповіді на питання по темі, завдання лабораторної роботи виконане з суттєвими недоліками;</p> <p>1 - студент погано розуміє розглянутий матеріал, завдання виконане неналежним чином;</p> <p>0 - студент зовсім не засвоїв розглянутий матеріал.</p> <p>Оцінювання змістових модулів (2 змістових модулів, 25 балів за кожний) відбувається шляхом написання студентом есе, тесту, написання програми, тощо.</p> <p>Бали оцінювання змістових модулів нараховуються за наступним співвідношенням:</p> <p>25-20 - розглянута тема відтворюється в повному обсязі, правильно, обґрунтовано, логічно, відповіді містять аналіз і систематизацію, аргументовані висновки. Засвідчено глибоке володіння матеріалом. Наведені приклади коду повністю робочі та відповідають темі. Можуть бути присутні несуттєві помилки та невідповідності;</p> <p>20-15 - відтворюється значна частина розглянутої теми. Виявлено знання і розуміння основних положень навчальної дисципліни, проте присутні неточності та/або невідповідності основній темі. Наведені приклади коду частково робочі, проте в загальному відповідають темі;</p> <p>15-10 - відстежується загальне розуміння розглянутої теми. Виявлені множинні неточності та невідповідності, пояснення наведеного коду відсутні, код функціонує із значними неточностями (або відсутні приклади запуску коду на виконання взагалі);</p> <p>10-5 – студент погано розуміє розглянуту тему. Виявлені суттєві неточності та невідповідності. Наведені приклади коду з суттєвими недоліками, або не відповідають темі;</p> <p>5 – 0 – студент взагалі не розуміє розглянуту тему. Тему не розкрито, кількість викладеного матеріалу не відповідає загальним нормам обраного виду роботи. Наведений код не робочий, або відсутній як такий.</p> <p>Критерії оцінювання результатів неформальної освіти: Нарахування балів відбувається за написання студентом тез доповідей на конференціях, наукових статей, участь у діяльності наукових гуртків, участь у наукових семінарах та круглих столах, конкурсах, участь у заходах неформальної освіти за отримання сертифікатів про проходження навчання на різних освітніх платформах (Coursera, Prometheus тощо), курсах на провідних ІТ компаніях за тематикою навчальної дисципліни. Кількість балів визначається відсотком покриття результатів відповідної активності до вимог результатів навчання з навчальної дисципліни.</p>
<p>Питання до заліку чи екзамену</p>	<ol style="list-style-type: none"> 1. Методи класифікації розпізнавання образів. 2. Матричні фільтри: згортка, фільтр розмиття, медіанний фільтр, ерозія і нарощування.

	<ol style="list-style-type: none"> 3. Скалярні і бінарні зображення. 4. Ідентифікація образів за кольором. 5. Зображення в частотній області. 2D перетворення Фур'є. 6. Класичні методи фільтрації зображень. 7. Матричні фільтри: згортка, фільтр розмиття, медіанний фільтр, ерозія і нарощування. Застосування згортки для аналізу зображень. 8. Фільтр Гаусса (ФНЧ). Фільтр Габора (ФВЧ). 9. Інтегральне представлення зображення. Контурний аналіз. 10. Виділення і опис контурів. Карти контурів. 11. Градієнт зображення. Прості детектори контурів. Оператор: Собеля, Лапласа, Кенні. 12. Прості детектори кутів зображення. Матриця Гессе. Детектор кутів Харріса і Стівенса, Алгоритм FAST. 13. Складні детектори контурів: лапласіан, гауссіан (LoG) і різниця гауссіанів (DoG). 14. Бінаризація зображення за пороговим значенням. 15. Шум в бінарних зображеннях. 16. Лінгвістика та її структура. Загальні поняття про мову. 17. Системи та мережі. Складні системи та мережі. Приклади. 18. Поняття статистичної лінгвістики. Методика вивчення лінгвістичної статистики. 19. Статичні та динамічні закони лінгвістики. 20. Механізми степеневих розподілів. 21. Пошук ключових слів у текстах. 22. Встановлення мови, авторства, стилістики та плагіату. 23. Мережеві властивості природних і рандомних текстів. 24. Класифікація прикладних лінгвістичних систем. 25. Інформаційні пошукові системи. Генерація тексту.
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

**Схема курсу “Методи та технології CV та NLP”
для студентів спеціальності 121 – Інженерія програмного забезпечення**

Тиж	Тема, план, короткі тези	Форма діяльності (заняття) **лекція, самостійна, дискусія, групова робота)	Література. Ресурси в Інтернеті	Завдання (лабораторна робота), год	Термін виконання
1	Основні поняття розпізнавання образів. Методи класифікації розпізнавання образів.	Лекція	1, 2,4 Сайт курсу	Вступне заняття. Техніка безпеки, перевірка обладнання та налаштування програмного забезпечення. Матричні фільтри: згортка, фільтр розмиття, медіанний фільтр, ерозія і нарощування.	2 тиж. сем

				щування.	
2	Скалярні і бінарні зображення. Векторні і RGB-зображення. Гістограми. Зміна контрасту зображення.	Лекція	1, 3, 4 Сайт курсу	Ідентифікація образів за кольором. Оцінка відмінностей гістограм кольору.	3 тиж. сем
3	Зображення в частотній області. 2D перетворення Фур'є. Аналіз спектра зображення. Фазово-конгруентна модель ознак зображення Локальне перетворення Фур'є в особливих точках зображення.	Лекція	1, 2, 4 Сайт курсу	Класифікація текстури об'єкта за допомогою параметрів гармонік в ряді Фур'є.	4 тиж. сем
4	Класичні методи фільтрації зображень. Матричні фільтри: згортка, фільтр розмиття, медіанний фільтр, ерозія і нарощування. Застосування згортки для аналізу зображень. Фільтр Гаусса (ФНЧ). Фільтр Габора (ФВЧ).	Лекція	2,3, 4 Сайт курсу	Класифікація текстури об'єкта за допомогою параметрів гармонік в ряді Фур'є.	5 тиж. сем
5	Інтегральне представлення зображення. Контурний аналіз. Виділення і опис контурів. Карти контурів. Градієнт зображення. Прості детектори контурів. Оператор: Собеля, Лапласа, Кенні. Прості детектори кутів зображення. Матриця Гессе. Детектор кутів Харріса і Стівенса, Алгоритм FAST. Складні детектори контурів: лапласіан гауссіан (LoG) і різниця гауссіанів (DoG)	Лекція	2, 3 Сайт курсу	Виділення і аналіз контурів. Метод пошуку границь Кенні.	6 тиж. сем
6	Бінаризація зображення за пороговим значенням. Алгоритми вибору порогового значення. Адаптивна бінаризація.	Лекція	1, 2, 3 Сайт курсу	Бінаризація зображення	7 тиж. сем
7	Шум в бінарних зображеннях. Методи математичної морфології.	Лекція	1, 2, 3 Сайт курсу	Змістовий модуль 1	8 тиж. сем
8	Комп'ютерна лінгвістика. Структура лінгвістики. Основні поняття Зв'язки комп'ютерної лінгвістики з галузями інформатики та систем штучного інтелекту. Лінгвістика та її структура. Загальні поняття про мову. Семантика. Прагматика.	Лекція	5, 6, 7 Сайт курсу	Попереднє опрацювання текстів. Дослідження швидкодії програм. Фонетичний і силабічний аналіз текстів східнослов'янськими мовами. Програмування задачі визначення співвідношення кількості голосних і	9 тиж. сем

				приголосних.	
9	Лінгвістика та системний підхід Системи та мережі. Складні системи та мережі. Приклади. Поняття статистичної лінгвістики. Методика вивчення лінгвістичної статистики.	Лекція	6, 7 Сайт курсу	Побудова гістограм словника текстової бази.	10 тиж. сем
10	Основні закони лінгвістики Статичні та динамічні закони лінгвістики. Статистика інших складних систем. Механізми ступеневих розподілів.	Лекція	5, 6, 7 Сайт курсу	Закони статистичної лінгвістики для окремих текстів. Статистика лексичних n-грам.	11 тиж. сем
11,12	Основи опрацювання природної мови Пошук ключових слів у текстах. Встановлення мови, авторства, стилістики та плагіату	Лекція	6, 7, 8 Сайт курсу	Визначення середньої довжини слів і речень.	13 тиж. сем
13,14	Скейлінг і мережеві властивості в лінгвістиці Скейлінг у лінгвістиці та інших складних системах. Мережеві властивості природних і рандомних текстів.	Лекція	5, 6, 7 Сайт курсу	Абсолютні методи визначення ключових слів у текстах. Метод TF-IDF. Визначення семантичного навантаження текстів на основі параметрів кластеризації.	15 тиж. сем
15,16	Огляд сучасних продуктів комп'ютерної лінгвістики Класифікація прикладних лінгвістичних систем. Мовна перевірка. Переклад. Інформаційні пошукові системи. Генерація тексту. Аналіз і синтез мовлення.	Лекція	6, 7, 8 Сайт курсу	Змістовий модуль 2 Підсумкове заняття.	16 тиж. сем