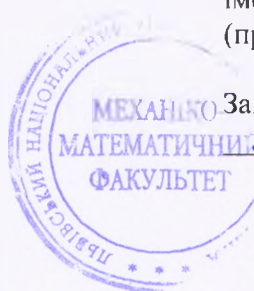


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Механіко-математичний факультет
Кафедра математичної статистики і диференціальних рівнянь

Затверджено

На засіданні кафедри математичної
статистики
і диференціальних рівнянь
факультету механіко-математичного
Львівського національного університету
імені Івана Франка
(протокол №11 від 22.06.2021 р.)



Завідувач кафедри Бугрій О.М.

Силабус з навчальної дисципліни
«ОБРОБКА ПРИРОДНОЇ МОВИ»,
що викладається в межах ОПП Статистика
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 112 Статистика

Львів

Назва дисципліни	Обробка природної мови
Адреса викладання дисципліни	Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1
Факультет, кафедра, за якою закріплена дисципліна	Механіко-математичний, Кафедра математичної статистики і диференціальних рівнянь
Галузь знань, шифр та назва спеціальності	11 «Математика та статистика» Спеціальність 112 «Статистика»
Викладачі дисципліни	Ярова Оксана Анатоліївна, кандидат фізико-математичних наук, асистент кафедри математичної статистики і диференціальних рівнянь
Контактна інформація викладачів	oksana.yarova@lnu.edu.ua кафедра математичної статистики і диференціальних рівнянь http://new.mmf.lnu.edu.ua/employee/yarova_o_a
Консультації з питань навчання по дисципліні відбуваються	Консультації за попередньою домовленістю. Для погодження часу консультацій слід писати на електронну пошту викладача.
Сторінка дисципліни	
Інформація про дисципліну	Дисципліна «Обробка природної мови» є нормативною дисципліною з спеціальності «112 Статистика» для освітньої програми «Статистичний аналіз даних», яка викладається в 6 семестрі в обсязі 3,5 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Обробка даних на природній мові передбачає вивчення комп'ютерної лінгвістики, починаючи від теорії математичного та лінгвістичного моделювання і до різноманітних технічних рішень.
Мета та цілі дисципліни	<i>Мета дисципліни:</i> ознайомити студентів з основними методами обробки природної мови; <i>Цілі дисципліни:</i> навчити аналізувати тексти та розробляти програми в Python.
Література для вивчення дисципліни	1. Natural Language Processing with Python. Steven Bird, Ewan Klein, and Edward Loper. 2. Liddy, E.D. 2001. Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker

	<p>3. Duan, Yucong; Cruz, Christophe (2011). "Formalizing Semantic of Natural Language through Conceptualization from Existence"</p> <p>4. Allen B. Downey, "Think Python – How to think like a computer scientist", 2nd ed., O'Reilly, 2016, p.291</p> <p>5. Mark Lutz, "Learning Python", 5th ed., O'Reilly, 2013, p.1594</p>
Обсяг курсу	48 годин аудиторних занять, з них 16 годин лекцій, 32 години лабораторних занять та 57 годин самостійної роботи
Очікувані результати навчання	<p>Після завершення цього курсу студент буде:</p> <p>Знати: методи обробки текстів та аналізу природної мови;</p> <p>Вміти: аналізувати тексти, структурувати інформацію та створювати програми в Python.</p>
Ключові слова	Python, NLP, аналіз тексту, класифікація тексту
Формат курсу	Очний
Теми	<p>Обробка мови в Python</p> <p>Текстові та лексичні ресурси</p> <p>Обробка вихідного тексту</p> <p>Нормалізація тексту та форматування</p> <p>Структуровані програми</p> <p>Категоризація та маркування слів</p> <p>Класифікація тексту</p> <p>Аналіз структури тексту</p>
Підсумковий контроль, форма	<p>Залік вкінці року.</p> <p>Залік – письмовий.</p>
Пререквізити	Для вивчення курсу студенти потребують базових знань з дисциплін: Інформатика та програмування, Теорія ймовірностей, Дискретна математика
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Презентації, лекції, практичні заняття
Необхідне обладнання	Комп'ютер
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться за 100-бальною шкалою.</p> <p>Бали нараховуються за наступним співвідношенням:</p> <ul style="list-style-type: none"> - Контрольні роботи: 50% семестрової оцінки; максимальна кількість балів 30; - Залік: 50% семестрової оцінки; максимальна

	<p>кількість балів 50. Загалом 100 балів.</p> <p>Академічна доброчесність: Очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в написанні програм є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та лабораторні зайняття курсу. Студенти повинні інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися термінів визначених для виконання всіх видів робіт та індивідуальних завдань, 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 тестуванні, самостійній роботі та бали підсумкового тестування. 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>Жодні форми порушення академічної доброчесності не толеруються.</p>
<p>Питання до екзамену (чи питання на контрольні роботи)</p>	<p>Обробка мови в Python Статистика мови Доступ до текстів Лексичні ресурси Послідовна схема NLP Нормалізація тексту</p>

	Форматування Розробка програм Категоризація та маркування слів Контролююча класифікація Дерево рішень Безконтекстна граматики Розробка граматики
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

**Схема курсу “Обробка природної мови”
для студентів спеціальності 112 – Статистика**

Тиж- ні	Лекційний курс		Лабораторні заняття		К-сть год сам. роб.
	Назва теми	К-сть год	Назва теми	К-сть год	
1	2	3	4	5	6
1	<i>Обробка мови в Python</i>	2	<i>Обробка мови в Python</i>	2	3
2			<i>Статистика мови</i>	2	3
3	<i>Текстові та лексичні ресурси</i>	2	<i>Доступ до текстів</i>	2	3
4			<i>Лексичні ресурси</i>	2	3
5	<i>Обробка вихідного тексту</i>	2	<i>Контрольна робота 1.</i>	2	3
6			<i>Послідовна схема NLP</i>	2	3
7	<i>Нормалізація тексту та форматування</i>	2	<i>Нормалізація тексту</i>	2	3
8			<i>Форматування</i>	2	4
9	<i>Структуровані програми</i>	2	<i>Розробка програм</i>	2	4
10			<i>Контрольна робота 2.</i>	2	4
11	<i>Категоризація та маркування слів</i>	2	<i>Категоризація та маркування слів</i>	2	4
12			<i>Контролююча класифікація</i>	2	4
13	<i>Класифікація тексту</i>	2	<i>Дерево рішень</i>	2	4
14			<i>Безконтекстна граматики</i>	2	4
15	<i>Аналіз структури тексту</i>	2	<i>Розробка граматики</i>	2	4
16			<i>Контрольна робота 3.</i>	2	4
	Викладач: Ярова О.А.	16	Викладач: Ярова О.А.	32	57