

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідас ДСТУ ISO 9001:2015	Ф-22.05- 05.01/035.00.1/Б/ОК29 -2023
	Екземпляр № 1	Арк 9 / 1

ЗАТВЕРДЖЕНО  
Вченою радою факультету  
інформаційно-комп'ютерних  
технологій

31.08. 2023 р., протокол № 5

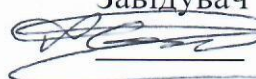
Голова Вченої ради  
Тетяна НІКІТЧУК



## РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ «Комп'ютерний аналіз текстових даних»

для здобувачів вищої освіти освітнього ступеня «бакалавр»  
спеціальності 035 «Філологія»  
освітньо-професійна програма  
«Філологія (прикладна лінгвістика)»  
факультет педагогічних технологій та освіти  
впродовж життя  
кафедра теоретичної та прикладної лінгвістики

Схвалено на засіданні  
кафедри комп'ютерної  
інженерії та кібербезпеки  
28.08 2023 р., протокол № 7

Завідувач кафедри  
 Андрій ЄФІМЕНКО

Гарант освітньо-професійної програми  
 Людмила МОГЕЛЬНИЦЬКА

Розробник: старший викладач кафедри комп'ютерної інженерії  
та кібербезпеки Оксана ОКУНЬКОВА, старший викладач  
кафедри комп'ютерних наук Галина МАРЧУК

Житомир  
2026 – 2027 н.р.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-22.05- 05.01/035.00.1/Б/ОК29 -2023
	Екземпляр № 1	Арк 9 / 2

## 1. Опис навчальної дисципліни

Найменування показників	Галузь знань, спеціальність, освітній ступінь	Характеристика навчальної дисципліни
		денна форма навчання
Кількість кредитів 3	Галузь знань 03 Гуманітарні науки	нормативна
Модулів – 3	Спеціальність: 035 «Філологія»	Рік підготовки:
Змістових модулів – 3		4-й
Загальна кількість годин - 90		Семестр
		7
Тижневих годин для денної форми навчання: аудиторних 3 самостійної роботи – 2,6	Освітній ступінь «бакалавр»	Лекції
		16 год.
		Практичні
		- год.
		Лабораторні
		32 год.
		Самостійна робота
		42 год.
		Вид контролю: екзамен

Співвідношення кількості годин аудиторних занять до самостійної та індивідуальної роботи становить:

для денної форми навчання –53 % аудиторних занять, 47 % самостійної та індивідуальної роботи.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-22.05- 05.01/035.00.1/Б/ОК29 -2023
	Екземпляр № 1	Арк 9 / 3

## 2. Мета та завдання навчальної дисципліни

**Метою навчальної дисципліни** «Комп'ютерний аналіз текстових даних» є засвоєння студентами принципів аналізу текстової інформації, отримання практичних навичок створення прикладних програм і реалізації алгоритмів обробки природної мови засобами мови Python.

**Завданнями вивчення навчальної дисципліни є:**

- ознайомити студентів з наявним програмним забезпеченням для аналізу текстової інформації;
- ознайомити студентів з функціоналом бібліотек Python для обробки природної мови;
- сформуванати навички написання програм для автоматичної нормалізації, векторизації та визначення тональності текстів;
- сформуванати навички розробки і написання комп'ютерних програм для аналізу природної мови.

Зміст навчальної дисципліни направлений на формування наступних **компетентностей**, визначених стандартом вищої освіти зі спеціальності 035 «Філологія», освітня програма «Філологія (прикладна лінгвістика)»:

**ФК 7.** Здатність до збирання й аналізу, систематизації та інтерпретації мовних, літературних, фольклорних фактів, інтерпретації та перекладу тексту (залежно від обраної спеціалізації).

**ФК 10.** Здатність здійснювати лінгвістичний, літературознавчий та спеціальний філологічний (залежно від обраної спеціалізації) аналіз текстів різних стилів і жанрів.

**ФК 13.** Здатність забезпечувати лінгвістичний супровід інформаційних систем, здійснювати автоматизовану обробку текстової інформації.

**ФК 14.** Здатність до використання мов програмування, інструментів тестування та оцінки якості програмного забезпечення лінгвістичного спрямування, технологій штучного інтелекту для розв'язання професійних завдань.

Отримані знання з навчальної дисципліни стануть складовими наступних **програмних результатів** навчання визначених стандартом вищої освіти зі спеціальності 35 «Філологія», освітня програма «Філологія (прикладна лінгвістика)»:

**ПРН 6.** Використовувати інформаційні й комунікаційні технології для вирішення складних спеціалізованих задач і проблем професійної діяльності.

**ПРН 12.** Аналізувати мовні одиниці, визначати їхню взаємодію та характеризувати мовні явища і процеси, що їх зумовлюють.

**ПРН 20.** Уміння створювати електронні мовні ресурси: корпуси, словники, перекладачі.

**ПРН 21.** Використовувати інструменти мов програмування, засоби й методи штучного інтелекту для здійснення автоматизованої обробки текстових даних.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-22.05- 05.01/035.00.1/Б/ОК29 -2023
	Екземпляр № 1	Арк 9 / 4

### **3. Програма навчальної дисципліни**

#### **Змістовий модуль 1. Аналіз текстової інформації. Text Mining.**

##### **Тема 1. Комп'ютерний аналіз текстових даних.**

Інформаційні технології в лінгвістичних дослідженнях. Аналіз текстів засобами офісних програм. Рецензування документів з використанням офісних програм та сервісів Google.

##### **Тема 2. Автоматизовані системи аналізу текстових даних.**

Огляд основних показників та інструментів SEO-аналізу тексту. Перевірка на відповідність вимогам пошукових систем. Методи проведення SEO-аналізу тексту. Особливості роботи алгоритмів у пошукових системах. Використання прикладних онлайн програм для проведення SEO – аналізу текстів.

**Тема 3. Автоматизовані системи реферування та анотування текстів.** Методи і прийоми автореферування текстів. Одномовне та багатомовне реферування. Алгоритми реферування. -Оцінка якості реферування. Використання прикладних програм узагальнення текстів.

#### **Змістовий модуль 2. Аналіз текстової інформації засобами мови Python.**

##### **Тема 4. Операції над символьними рядками.**

Символьні рядки. Тип даних str. Екрановані послідовності. Витяг символів. Текстові та CSV-файли, створення, запис та читання. Візуалізація лінгвістичних досліджень. Регулярні вирази.

#### **Змістовий модуль 3. Аналіз текстової інформації засобами бібліотеки NLTK мови Python.**

##### **Тема 5. Задачі аналізу текстової інформації засобами бібліотеки NLTK.**

Постановка задачі аналізу текстової інформації. Етапи аналізу текстів. Добування ключових понять із тексту. Класифікація текстових документів. Методи кластеризації текстових документів. Засоби аналізу текстової інформації. Токенізація, нормалізація текстів. Визначення тональності.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-22.05- 05.01/035.00.1/Б/ОК29 -2023
	Екземпляр № 1	Арк 9 / 5

#### 4. Структура (тематичний план) навчальної дисципліни

Змістові модулі і теми	Кількість годин			
	денна форма			
	усь ого	лекц ії	лабо рато рні	с. р.
<b>Модуль 1</b>				
<b>Змістовий модуль 1. Аналіз текстової інформації. Text Mining.</b>				
Тема 1. Аналіз текстів засобами офісних програм та сервісів Google.	8	2	4	2
Тема 2. Методи та засоби проведення SEO-аналізу тексту.	12	2	4	6
Тема 3. Автоматизовані системи реферування текстів.	12	2	4	6
<i><b>Разом за змістовий модуль 1</b></i>	<b>32</b>	6	12	14
<b>Модуль 2</b>				
<b>Змістовий модуль 2. Аналіз текстової інформації засобами мови Python.</b>				
Тема 4. Операції над символьними рядками. Регулярні вирази	14	2	4	8
<i><b>Разом за змістовий модуль 2</b></i>	<b>14</b>	2	4	8
<b>Модуль 3</b>				
<b>Змістовий модуль 3. Аналіз текстової інформації засобами бібліотеки NLTK мови Python.</b>				
Тема 5. Задачі аналізу текстової інформації засобами бібліотеки NLTK.	44	8	16	20
<i><b>Разом за змістовий модуль 3</b></i>	<b>44</b>	8	16	20
<b>РАЗОМ:</b>	<b>90</b>	16	32	42

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-22.05- 05.01/035.00.1/Б/ОК29 -2023
	Екземпляр № 1	Арк 9 / 6

## 5. Темы лабораторних занять

№ з/п	Назва теми	Кількість годин
		денна форма
1.	Рецензування текстової інформації в офісних програмах та хмарних сервісів.	4
2.	Дослідження програмного забезпечення для SEO-аналізу текстових даних.	4
3.	Дослідження програмного забезпечення реферування (узагальнення) текстів.	4
4.	Засоби мови Python обробки текстових даних. Регулярні вирази.	4
5.	Бібліотека NLTK мови Python. Токенізація текстів. Стемінг та лемітизація тексту.	4
6.	Використання нормалізованих даних для аналізу тексту. Векторне представлення. Метод TF/ IDF.	4
7.	Використання n- грам для проведення аналізу текстів. Візуалізація результатів.	4
8.	Аналіз тональності текстів.	4
Разом		32

## 6. Завдання для самостійної роботи

- Тема 1. Історія та перспективи розвитку обробки природної мови. Програмне забезпечення інтелектуальних систем. Розробка природномовних інтерфейсів. Колективне рецензування документів.
- Тема 2. Основні пункти SEO аналізу текстів. Особливості роботи алгоритмів. Перефразування. Карти та хмари слів.
- Тема 3. Алгоритми реферування. Оцінка якості реферування
- Тема 4. Методи роботи з символічними рядками. Візуалізація лінгвістичних досліджень. Використання регулярних виразів. Файли
- Тема 5. Класифікація текстових документів. Методи кластеризації текстових документів. Бібліотека NLTK. Засоби аналізу текстової інформації засобами бібліотеки SpaCy.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-22.05- 05.01/035.00.1/Б/ОК29 -2023
	Екземпляр № 1	Арк 9 / 7

## 7. Методи навчання

У процесі викладання даної дисципліни використовуються наступні методи навчання:

- вербальні (лекції, пояснення, розповідь, бесіда, інструктаж);
- наочні (ілюстрація, демонстрація);
- практичні (різні види вправ та завдань, виконання лабораторних робіт);
- пояснювально-ілюстративний (передбачає пред'явлення готової інформації викладачем та її засвоєння студентами);
- репродуктивний, в основу якого покладено виконання різного роду завдань за зразком;
- метод проблемного викладу;
- частково-пошуковий (евристичний);
- дискусійний метод.

## 8. Методи контролю

- усне опитування; тестовий контроль;
- виконання практичних завдань (в тому числі, у цифровому освітньому середовищі);
- перевірка домашнього завдання (задачі, вправи, окремі розрахунки)
- проведення контрольних-модульних робіт;
- захист звітів з лабораторних робіт
- екзамен

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-22.05- 05.01/035.00.1/Б/ОК29 -2023
	Екземпляр № 1	Арк 9 / 8

## 9. Розподіл балів

Модулі та їх елементи	Форма контролю	Максимальна кількість балів
<b>Разом за змістовий модуль 1</b>		
<b>Змістовий модуль 2. Аналіз текстової інформації. Text Mining.</b>		
Лабораторна робота № 1	Виконання і захист ЛР	10
Лабораторна робота № 2	Виконання і захист ЛР	10
Лабораторна робота № 3	Виконання і захист ЛР	10
<b>Разом за змістовий модуль 2</b>		<b>30</b>
<b>Змістовий модуль 3. Аналіз текстової інформації засобами мови Python.</b>		
Лабораторна робота № 4	Виконання і захист ЛР	10
<b>Разом за змістовий модуль 2</b>		<b>10</b>
<b>Змістовий модуль 3. Аналіз текстової інформації засобами бібліотеки NLTK мови Python.</b>		
Лабораторна робота № 5	Виконання і захист ЛР	10
Лабораторна робота № 6	Виконання і захист ЛР	10
Лабораторні роботи № 7	Виконання і захист ЛР	10
Лабораторні роботи № 8	Виконання і захист ЛР	10
<b>Разом за змістовий модуль 3</b>		<b>40</b>
<b>Тестовий контроль</b>		
Тест 1		10
Тест 2		10
<b>Разом тести</b>		<b>20</b>
<b>Оцінка по дисципліні</b>		<b>100</b>

## Шкала оцінювання

За шкалою	Екзамен	Бали
A	Відмінно	90-100
B	Добре	82-89
C		74-81
D	Задовільно	64-73
E		60-63
FX	Незадовільно	35-59
F		0-34



Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-22.05- 05.01/035.00.1/Б/ОК29 -2023
	Екземпляр № 1	Арк 9 / 9

## 10. Рекомендована література

### Основна література

1. Mathangi Sri Practical Natural Language Processing with Python/ - Apress, 2021 - 253 с.
2. Tunstall Lewis, Leandro von Werra Natural Language Processing with Transformers. Revised Edition. - O'Reilly Media, 2022 – 406 с.
3. Костюченко А.О. Основи програмування мовою Python: навчальний посібник. Ч.: ФОП Баликіна С.М., 2020. -180 с.

### Допоміжна література

1. Васильєв О. М. Програмування мовою Python. Тернопіль: Навчальна книга – Богдан, 2019. – 504с.
2. Руденко В., Жугастров О. Інформатика. Основи алгоритмізації та програмування мовою Python. Харків: Ранок, 2019. – 192 с.
3. Копей В. Б. Мова програмування Python для інженерів і науковців : навч. посіб. / В. Б. Копей. - Івано-Франківськ : ІФНТУНГ, 2019. - 272 с.
4. Король В. Я., Окунькова О. О., Спрощений та розширений підходи обробки природної мови. Тези V Всеукраїнської науково-технічної конференції «Комп'ютерні технології: інновації, проблеми, рішення» 01-02 грудня 2022 року. Житомир : «Житомирська політехніка», 2022. С.172.
5. Марчук Г.В., Левківський В.Л., Каліберда С.С. Інтелектуальний аналіз даних. Біоніка інтелекту. Харків: ХНУРЕ, 2019. №1 (92). – С. 65–70.
6. Марчук Г.В., Левківський В.Л., Марчук Д.К., Муковоз В.С. Дослідження і аналіз можливостей чат-боту зі штучним інтелектом ChatGPT. Вчені записки Таврійського національного університету імені В.І. Вернадського Серія: Технічні науки Том 34 (73) № 4 2023. С. 71-78.
7. Окунькова О.О. Огляд безкоштовних сервісів автореферування текстів. Проблеми та перспективи розвитку науки, освіти і суспільства в ХХІ столітті: збірник тез доповідей міжнародної науково-практичної конференції (Рівне, 27 квітня 2023 р.). Рівне: ЦФЕНД, 2023. С. 60-61
8. Окунькова О.О. Сучасні інформаційні технології аналізу україномовних текстів. Вісник Кременчуцького національного університету імені Михайла Остроградського. Серія: Комп'ютерні науки – Кременчук: КрНУ, 2023. – Випуск 1(138). С. 73-79

### 11. Інформаційні ресурси в Інтернеті

1. Основи для письма «все в одному» <https://smodin.io/uk>
2. Лінгвістичний інструментарій <https://linguakit.com/>
3. Site-Analyzer <https://site-analyzer.pro/uk/services-seo/text-semantic/>
4. Edu-Birde <https://edubirdie.com/>
5. Словник NLP <http://surl.li/pvkku>
6. NLP <http://surl.li/pvkly>
7. Документація <https://www.nltk.org/>
8. CS50: Вступ до штучного інтелекту з Python [https://prometheus.org.ua/course/course-v1:HarvardUniversity+CS50\\_AI101+2023\\_T3](https://prometheus.org.ua/course/course-v1:HarvardUniversity+CS50_AI101+2023_T3)