

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРЬСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 1

ЗАТВЕРДЖЕНО

Вченою радою факультету
педагогічних технологій та освіти
впродовж життя

27 серпня 2024 р., протокол № 7

Голова Вченої ради

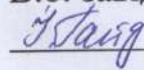
Оксана ЧЕРНИШ

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ «Текст Майнінг: інтелектуальний аналіз тексту»

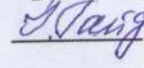
для здобувачів вищої освіти освітнього ступеня «бакалавр»
спеціальності 014 «Середня освіта»
спеціалізація: 014.021 «Англійська мова і література»
освітньо-професійна програма «Англійська мова і література»
факультет педагогічних технологій та освіти впродовж життя
кафедра педагогічних технологій та мовної підготовки

Схвалено на засіданні кафедри
педагогічних технологій та мовної
підготовки

26 серпня 2024 р.,
протокол № 8

В.о. завідувача кафедри
 Ірина ГАЙДАЙ

Гарант освітньо-професійної
програми

 Ірина ГАЙДАЙ

Розробник: д.філол.н., професор кафедри теоретичної та прикладної лінгвістики
КАПРАНОВ Ян

Житомир
2024

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 2

Робоча програма навчальної дисципліни «Текст Майнинг: інтелектуальний аналіз тексту» для здобувачів вищої освіти освітнього ступеня «бакалавр» спеціальності 014 «Середня освіта», спеціалізація: 014.021 «Англійська мова і література», освітньо-професійна програма «Англійська мова і література» затверджена Вченою радою факультету педагогічних технологій та освіти впродовж життя від 27 серпня 2024 р., протокол № 7.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/3

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, спеціальність, освітній ступінь	Характеристика навчальної дисципліни	
		денна форма навчання	заочна форма навчання
Кількість кредитів 3	Галузь знань 01 «Освіта/Педагогіка»	обов'язкова (обов'язкова, вибіркова)	
Модулів – 2	Спеціальність 014 «Середня освіта» Спеціалізація: 014.021 «Англійська мова і література»	Рік підготовки:	
Змістових модулів – 2		4-й	—
		Семестр	
Загальна кількість годин – 90		8	
		Лекції	
Тижневих годин для денної форми навчання: Аудиторних – 3 самостійної роботи – 4,5	Освітній ступінь «бакалавр»	12 год.	— год.
		Практичні	
		24 год.	— год.
		Лабораторні	
		— год.	— год.
		Самостійна робота	
		54 год.	— год.
		Вид контролю: залік	

Частка аудиторних занять і частка самостійної та індивідуальної роботи у загальному обсязі годин з навчальної дисципліни становить:

для денної форми навчання – 40% аудиторних занять, 60% самостійної та індивідуальної роботи.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 4

2. Мета та завдання навчальної дисципліни

Метою вивчення навчальної дисципліни «Текст майнінг: інтелектуальний аналіз тексту» є ознайомлення студентів з технологією Текст майнінг, метою якої є отримання інформації з колекцій текстових документів, ґрунтуючись на застосуванні ефективних, у практичному плані, методів машинного навчання та обробки природної мови. Інтелектуальний аналіз тексту використовує всі ті ж підходи до перероблювання інформації, що й інтелектуальний аналіз даних, однак різниця між цими напрямками проявляється лише в кінцевих методах, а також у тому, що інтелектуальний аналіз даних має справу зі сховищами та базами даних, а не електронними бібліотеками та корпусами текстів. Докладно розглядаються методи, інструментальні засоби й застосування Текст Майнінг. Опис кожного методу супроводжується конкретним прикладом його використання; застосування набутих знань і вмінь для вирішення питань міжкультурної комунікації та професійних потреб.

Завданнями навчальної дисципліни «Текст майнінг: інтелектуальний аналіз тексту» є:

- формування у студентів базових знань про основні поняття та методи текст майнінгу, зокрема обробки природної мови;
- вивчення алгоритмів та підходів до очищення, попередньої обробки та трансформації текстових даних.
- вдосконалення мовленнєвої підготовки завдяки використанню автентичних англомовних матеріалів;
- використання методів видобування інформації з тексту, такі як виявлення ключових слів, класифікація, кластеризація, аналіз тональності та тематичне моделювання.;
- підготовка висококваліфікованих фахівців з технологій збору та аналізу текстових даних із соціальних мереж, веб-ресурсів та інших джерел.

Зміст навчальної дисципліни направлений на формування наступних **компетентностей**, визначених стандартом вищої освіти зі спеціальності 014 «Середня освіта», спеціалізація: 014.021 «Англійська мова і література», освітньо-професійна програма «Англійська мова і література».

Загальні компетентності (ЗК)

ЗК-9. Здатність використовувати знання іноземної мови в освітній діяльності.

ЗК-12. Здатність проведення досліджень на відповідному рівні.

Спеціальні (фахові, предметні) компетентності (СК)

СК-10. Здатність інтерпретувати й зіставляти мовні та літературні явища, використовувати різні методи й методики аналізу тексту.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/5

СК-16. Здатність розуміти вимоги до 7 діяльності за спеціальністю, зумовлені необхідністю забезпечення сталого розвитку України, її зміцнення як демократичної, соціальної, правової держави.

Отримані знання з навчальної дисципліни стануть складовими наступних програмних **результатів навчання** за спеціальністю 014 «Середня освіта», спеціалізація: 014.021 «Англійська мова і література», освітньо-професійна програма «Англійська мова і література»:

Результати навчання (РН), спільні для всіх предметних спеціальностей

РН-2. Обізнаний із елементами теоретичного й експериментального (пробного) дослідження в професійній сфері та методами їхньої реалізації.

РН-9. Володіє різними видами аналізу художнього твору, визначає його жанрово-стильову своєрідність, місце в літературному процесі, традиції й новаторство, зв'язок твору із фольклором, міфологією, релігією, філософією, значення для національної та світової культури.

РН-11. Порівнює мовні та літературні факти, явища, визначає їхні подібності й відмінності.

РН-13. Володіє основами професійної культури, має здатність створювати й редагувати тексти професійного змісту державною та іноземною мовою.

Під час вивчення навчальної дисципліни здобувачі вищої освіти зможуть отримати наступні Soft skills:

- *комунікативні навички*: письмове, вербальне й невербальне спілкування; вести дискусію і відстоювати свою позицію; навички працювати в команді;
- *уміння виступати привселюдно*: навички, необхідні для виступів на публіці; навички проведення презентації;
- *керування часом*: уміння справлятися із завданнями вчасно;
- *гнучкість і адаптивність*: гнучкість, адаптивність і здатність змінюватися; уміння аналізувати ситуацію, орієнтування на вирішення проблеми;
- *лідерські якості*: уміння спокійно працювати в напруженому середовищі; уміння ухвалювати рішення; уміння ставити мету, планувати діяльність;
- *особисті якості*: креативне й критичне мислення; етичність, чесність, терпіння, повага до оточуючих.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 6

3. Програма навчальної дисципліни

МОДУЛЬ 1

Змістовий модуль 1. ОСНОВИ ТЕКСТ МАЙНІНГУ

Тема 1. Вступ до текст майнінгу (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Основні поняття та задачі інтелектуального аналізу тексту.
2. Відмінності між структурованими та неструктурованими даними.
3. Лінгвістичні особливості текстів для аналізу.

Тема 2. Інструменти та середовища для текст майнінгу (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Огляд інструментів (Python, NLTK, spaCy).
2. Основи роботи з текстовими даними у Python.
3. Формати зберігання тексту (txt, csv, JSON).

Тема 3. Попередня обробка тексту (Preprocessing) (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Лемматизація та стемінг.
2. Очистка тексту: видалення стоп-слів, пунктуації, нормалізація.
3. Токенізація: слово, речення, абзац.

Тема 4. Статистичний аналіз тексту (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Частотний аналіз слів і фраз.
2. Біграми та n-грам моделі.
3. Хмари слів для візуалізації частотності.

Тема 5. Виділення ключових слів і інформації (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Методи TF-IDF та TextRank.
2. Автоматизоване узагальнення текстів.
3. Видобування іменованих сутностей (Named Entity Recognition, NER).

Тема 6. Аналіз тональності тексту (Sentiment Analysis) (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 7

1. Основи емоційного аналізу тексту.
2. Словникові та машинні методи.
3. Аналіз тональності у відгуках та соціальних мережах.

Тема 7. Тематичне моделювання (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Алгоритми тематичного аналізу (LDA, NMF).
2. Побудова та інтерпретація моделей.
3. Використання тематичного моделювання для класифікації текстів.

Тема 8. Класифікація текстів (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Основи класифікації текстів: ручна розмітка та навчання моделі.
2. Алгоритми: Наївний Байєс, SVM, Random Forest.
3. Метрики оцінювання (accuracy, precision, recall, F1-score).

МОДУЛЬ 2

Змістовий модуль 2. СУЧАСНІ ТЕХНОЛОГІЇ ТЕКСТ МАЙНИНГУ

Тема 1. Векторизація тексту (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Методи векторизації: Bag of Words, TF-IDF.
2. Вступ до Word Embeddings (Word2Vec, FastText).
3. Порівняння методів векторизації.

Тема 2. Аналіз семантичної схожості текстів (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Моделі Sentence Embeddings (Sentence-BERT, USE).
2. Вимірювання семантичної схожості.
3. Приклади використання: кластеризація, пошук схожих документів.

Тема 3. Глибоке навчання у текст майнінгу (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Основи роботи з моделями BERT і GPT.
2. Файнтюнінг моделей для задач класифікації та генерації тексту.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 8

3. Застосування трансформерів у текстовому аналізі.

Тема 4. Автоматична генерація тексту (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Методи генерації тексту: LSTM, GPT.
2. Приклади створення текстів: статті, резюме, творчі тексти.
3. Етичні аспекти автоматичної генерації текстів.

Тема 5. Аналіз текстів соціальних мереж (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Збір даних із Twitter, Facebook, Reddit.
2. Аналіз емоцій та популярності постів.
3. Визначення трендів на основі текстового аналізу.

Тема 6. Розробка чат-ботів (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Основи роботи чат-ботів.
2. Інтеграція NLU (Natural Language Understanding).
3. Створення прототипу чат-бота з базовою логікою.

Тема 7. Текстовий аналіз у різних галузях (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Аналіз юридичних текстів.
2. Аналіз медичних текстів (лікарські звіти, рецепти).
3. Аналіз наукових статей та автоматичне створення рефератів.

Тема 8. Етичні аспекти та виклики текстового аналізу (ЗК-9, ЗК-12, СК-10, СК-16, РН2, РН9, РН11, РН13)

1. Конфіденційність і етичні аспекти роботи з текстовими даними.
2. Виклики та ризики автоматичного аналізу тексту.
3. Майбутні перспективи текст майнінгу для лінгвістів.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/9

4. Структура (тематичний план) навчальної дисципліни

Змістові модулі і теми	Кількість годин							
	денна форма				заочна форма			
	усього	лекції	практичні	самостійна робота	усього	лекції	практичні	самостійна робота
МОДУЛЬ 1								
Змістовий модуль 1. ОСНОВИ ТЕКСТ МАЙНІНГУ								
Тема 1. Вступ до текст майнінгу	5	2	-	3	-	-	-	-
Тема 2. Інструменти та середовища для текст майнінгу	5	-	2	3	-	-	-	-
Тема 3. Попередня обробка тексту (Preprocessing)	5	-	2	3	-	-	-	-
Тема 4. Статистичний аналіз тексту	7	2	2	3	-	-	-	-
Тема 5. Виділення ключових слів і інформації	5	-	2	3	-	-	-	-
Тема 6. Аналіз тональності тексту (Sentiment Analysis)	5	-	2	3	-	-	-	-
Тема 7. Тематичне моделювання	5	2	-	3	-	-	-	-
Тема 8. Класифікація текстів	5	-	2	3	-	-	-	-
Разом за змістовий модуль 1	42	6	12	24	-	-	-	-
РАЗОМ ЗА МОДУЛЬ 1	42	6	12	24	-	-	-	-
МОДУЛЬ 2								
Змістовий модуль 2. СУЧАСНІ ТЕХНОЛОГІЇ ТЕКСТ МАЙНІНГУ								
Тема 1. Векторизація тексту	7	-	2	4	-	-	-	-
Тема 2. Аналіз семантичної схожості текстів	8	2	-	4	-	-	-	-
Тема 3. Глибоке навчання у текст майнінгу	7	-	2	4	-	-	-	-
Тема 4. Автоматична генерація тексту.	8	2	2	4	-	-	-	-
Тема 5. Аналіз текстів соціальних мереж	7	-	2	4	-	-	-	-
Тема 6. Розробка чат-ботів	8	-	2	4	-	-	-	-
Тема 7. Текстовий аналіз у різних галузях	7	2	-	3	-	-	-	-
Тема 8. Етичні аспекти та виклики	7	-	2	3	-	-	-	-

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/10

Змістові модулі і теми	Кількість годин							
	денна форма				заочна форма			
	усього	лекції	практичні	самостійна робота	усього	лекції	практичні	самостійна робота
текстового аналізу								
<i>Разом за змістовий модуль 2</i>	48	6	12	30	-	-	-	-
РАЗОМ ЗА МОДУЛЬ 2	48	6	12	30	-	-	-	-
ВСЬОГО	90	12	24	54	-	-	-	-

5. Теми практичних занять

№ з/п	Назва теми	Кількість годин	
		денна форма	заочна форма
МОДУЛЬ 1			
Змістовий модуль 1. ОСНОВИ ТЕКСТ МАЙНІНГУ			
1	Вступ до текст майнінгу	-	-
2	Інструменти та середовища для текст майнінгу	2	-
3	Попередня обробка тексту (Preprocessing)	2	-
4	Статистичний аналіз тексту	2	-
5	Виділення ключових слів і інформації	2	-
6	Аналіз тональності тексту (Sentiment Analysis)	2	-
7	Тематичне моделювання	-	-
8	Класифікація текстів	2	-
РАЗОМ ЗА МОДУЛЬ 1		12	-
МОДУЛЬ 2			
Змістовий модуль 1. СУЧАСНІ ТЕХНОЛОГІЇ ТЕКСТ МАЙНІНГУ			
1	Векторизація тексту	2	-
2	Аналіз семантичної схожості текстів	-	-
3	Глибоке навчання у текст майнінгу	2	-
4	Автоматична генерація тексту.	2	-
5	Аналіз текстів соціальних мереж	2	-
6	Розробка чат-ботів	2	-
7	Текстовий аналіз у різних галузях	-	-
8	Етичні аспекти та виклики текстового аналізу	2	-

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 11

№ з/п	Назва теми	Кількість годин	
		денна форма	заочна форма
РАЗОМ ЗА МОДУЛЬ 2		12	-
РАЗОМ		24	-

6. Завдання для самостійної роботи

№ з/п	Назва теми	Кількість годин	
		денна форма	заочна форма
МОДУЛЬ 1			
Змістовий модуль 1. ОСНОВИ ТЕКСТ МАЙНІНГУ			
1	Вступ до текст майнінгу - Знайдіть три приклади практичного застосування текст майнінгу в різних галузях (наприклад, медицина, маркетинг, соціальні мережі). - Сформулюйте задачі текстового аналізу для кожного прикладу. - Опишіть, у чому полягає відмінність між структурованими, напівструктурованими та неструктурованими даними. Наведіть приклади для кожної категорії.	3	-
2	Інструменти та середовища для текст майнінгу - Встановіть Python і бібліотеки для текстового аналізу (NLTK, spaCy, Scikit-learn, Pandas). Напишіть короткий звіт із встановлення та виконання тестових скриптів. - Підготуйте текстовий файл із набором статей або оглядів. Завантажте файл у Python і проаналізуйте його структуру (кількість символів, слів, рядків).	3	-
3	Попередня обробка тексту (Preprocessing) - Візьміть текстовий документ (наприклад, новинну статтю або текст з інтернету) і виконайте: - Видалення пунктуації. - Приведення до нижнього регістру. - Видалення стоп-слів.	3	-

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 12

	Виконайте лемматизацію або стемінг для заданого тексту. Порівняйте результати: які слова було змінено та які залишилися без змін.		
4	Статистичний аналіз тексту - Підрахуйте частоту всіх слів у вибраному тексті. Виведіть 10 найчастотніших слів. - Побудуйте хмару слів для даного тексту. Оцініть, які слова найбільш характерні для цього тексту.	3	-
5	Виділення ключових слів і інформації - Використовуючи метод TF-IDF, виділіть ключові слова для заданого текстового документа. Наведіть топ-5 слів із найвищими значеннями TF-IDF. - Реалізуйте видобування іменованих сутностей (NER) із тексту за допомогою бібліотеки spaCy. Вкажіть кількість імен, місць і організацій, які було виявлено.	3	-
6	Аналіз тональності тексту (Sentiment Analysis) - Виконайте аналіз тональності для набору відгуків (наприклад, про товар чи фільм) за допомогою TextBlob або VADER. Визначте кількість позитивних, негативних і нейтральних відгуків. - Порівняйте результати аналізу тональності, отримані різними методами (наприклад, словниковий метод VADER і модель BERT). Які розбіжності ви помітили?	3	-
7	Тематичне моделювання - Виконайте тематичне моделювання для невеликого корпусу текстів за допомогою LDA. Визначте 2–3 основні теми та наведіть приклади слів, характерних для кожної теми. - Застосуйте тематичне моделювання для текстів новин. Які теми були виділені? Як ви їх інтерпретуєте?	3	-
8	Класифікація текстів Візьміть невеликий набір текстів, розподілений за категоріями (наприклад, "Наука", "Спорт",	3	-

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 13

	"Політика"). Реалізуйте класифікацію текстів за допомогою Наївний Байєсівського класифікатора. Підготуйте навчальний та тестовий набори даних для класифікації текстів. Визначте точність моделі, використовуючи метрики precision, recall та F1-score.		
РАЗОМ ЗА МОДУЛЬ 1		24	-
МОДУЛЬ 2			
Змістовий модуль 2. СУЧАСНІ ТЕХНОЛОГІЇ ТЕКСТ МАЙНІНГУ			
1	Векторизація тексту - Векторизуйте заданий текстовий корпус за допомогою методів Bag of Words і TF-IDF. - Порівняйте отримані результати (зокрема, кількість ненульових елементів у матриці). - Використовуючи Word2Vec або FastText, створіть векторне представлення слів із заданого тексту. - Обчисліть косинусну подібність між векторами слів "освіта" і "навчання".	4	-
2	Аналіз семантичної схожості текстів - Підготуйте кілька пар текстів і обчисліть їх семантичну схожість за допомогою Sentence-BERT. - Поясніть, які пари текстів виявилися найбільш схожими та чому. - Реалізуйте завдання кластеризації текстів (наприклад, новинні статті) на основі їх векторних представлень. Візуалізуйте результати кластеризації.	4	-
3	Глибоке навчання у текст майнінгу - Використовуючи модель BERT, виконайте класифікацію текстів (наприклад, розподіл на позитивні й негативні відгуки). - Порівняйте точність BERT із простішими моделями (наприклад, Logistic Regression). - Навчіть модель для завдання NER (Named Entity Recognition) на основі глибокого навчання. Визначте якість моделі за допомогою метрик precision і recall.	4	-

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 14

4	Автоматична генерація тексту. - Використовуючи модель GPT, згенеруйте текст на задану тему (наприклад, "Технології майбутнього"). - Проаналізуйте логічність і лексичну різноманітність створеного тексту. - Змініть налаштування температури у GPT і згенеруйте тексти для одного й того ж запиту. Порівняйте, як змінюється якість тексту залежно від параметра.	4	-
5	Аналіз текстів соціальних мереж - Зберіть невелику вибірку текстів із Twitter або Facebook за певним хештегом чи ключовим словом. - Виконайте аналіз тональності цих текстів (позитивні, негативні, нейтральні). - Визначте найбільш популярні теми у зібраних текстах за допомогою тематичного моделювання (LDA).	4	-
6	Розробка чат-ботів - Створіть базовий чат-бот із фіксованими відповідями на типові запити (наприклад, питання про навчання). - Розширте функціонал чат-бота, додавши можливість аналізу ключових слів у запиті. - Інтегруйте в чат-бот NLU (наприклад, Rasa або Dialogflow) для розпізнавання інтенцій. Наведіть приклади кількох сценаріїв роботи бота.	4	-
7	Текстовий аналіз у різних галузях - Виконайте аналіз тексту з певної галузі (наприклад, медичні звіти або юридичні документи). Виділіть ключові слова та теми. - Проаналізуйте набори відгуків про продукт або послугу. Визначте основні скарги та пропозиції клієнтів.	3	-
8	Етичні аспекти та виклики текстового аналізу Проведіть аналіз можливих етичних ризиків при використанні текст майнінгу у вашому проєкті (наприклад, робота з конфіденційними даними чи соціальними мережами).	3	-

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 15

	Напишіть есе на тему "Як уникнути упередженості в моделях текстового аналізу?" з аналізом практичних прикладів.		
РАЗОМ ЗА МОДУЛЬ 2		30	-
РАЗОМ		54	-

7. Індивідуальні самостійні завдання

Індивідуальні завдання є однією з форм навчального процесу, що передбачає створення умов для найповнішої реалізації творчості можливостей студентів і має на меті поглиблення, узагальнення та закріплення знань, які студенти можуть одержати в процесі вивчення курсу даної дисципліни, а також застосування цих знань на практиці.

Ці завдання сприяють розвитку навичок самостійної роботи, критичного мислення, а також покращують мовну компетенцію студентів. Індивідуальні завдання включають:

- підготовку письмових есе на задану тему;
- виконання тематичних перекладів текстів із англійської мови на рідну та навпаки;
- підготовку усних презентацій на актуальні теми з курсу;
- розробку проєктів, що демонструють застосування набутих знань на практиці;
- аналітичне читання наукових статей та їх критичний огляд;
- виконання завдань з лексичних та граматичних тем, що виходять за межі основної програми курсу.

Індивідуальні завдання виконуються за попереднім погодженням із викладачем і враховують рівень знань студента, а також його особисті інтереси у вивченні англійської мови.

8. Методи навчання

Під час викладання навчальної дисципліни використовуються методи навчання, що сприяють досягненню відповідних програмних результатів.

Результат навчання	Методи навчання
РН-2. Обізнаний із елементами теоретичного й експериментального (пробного) дослідження в професійній сфері та методами їхньої реалізації.	– Вербальні методи (пояснення, розповідь, бесіда) – Наочні методи (презентація, демонстрація відеоматеріалів)

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 16

Результат навчання	Методи навчання
	– Практичні методи (виконання різних видів вправ, практичних завдань, творчі завдання) – Дискусійний метод (практикувати мову в ситуаціях, близьких до реальних професійних ситуацій) – Метод активного навчання (проведення ділових ігор, мозковий штурм, командна робота) – Методи самостійної роботи (анотування опрацьованого матеріалу, написання есе, підготовка тез доповідей) – Метод кейс-стаді (аналіз реальних або змодельованих ситуацій, що дозволяє студентам застосовувати набуті знання у вирішенні практичних завдань) – Метод проектів (виконання студентами групових або індивідуальних проектів на основі професійних потреб)
РН-9. Володіє різними видами аналізу художнього твору, визначає його жанрово-стильову своєрідність, місце в літературному процесі, традиції й новаторство, зв'язок твору із фольклором, міфологією, релігією, філософією, значення для національної та світової культури.	– Вербальні методи (пояснення, розповідь, бесіда) – Наочні методи (презентація, демонстрація відеоматеріалів) – Практичні методи (виконання різних видів вправ, практичних завдань, творчі завдання) – Дискусійний метод (практикувати мову в ситуаціях, близьких до реальних професійних ситуацій) – Метод активного навчання (проведення ділових ігор, мозковий штурм, командна робота) – Методи самостійної роботи (анотування опрацьованого матеріалу, написання есе, підготовка тез доповідей) – Метод кейс-стаді (аналіз реальних або змодельованих ситуацій, що дозволяє студентам застосовувати набуті знання у вирішенні практичних завдань) – Метод проектів (виконання студентами групових або індивідуальних проектів на основі професійних потреб)
РН-11. Порівнює мовні та літературні факти, явища, визначає їхні подібності й відмінності.	– Вербальні методи (пояснення, розповідь, бесіда) – Наочні методи (презентація, демонстрація відеоматеріалів)

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 17

Результат навчання	Методи навчання
	– Практичні методи (виконання різних видів вправ, практичних завдань, творчі завдання) – Дискусійний метод (практикувати мову в ситуаціях, близьких до реальних професійних ситуацій) – Метод активного навчання (проведення ділових ігор, мозковий штурм, командна робота) – Методи самостійної роботи (анотування опрацьованого матеріалу, написання есе, підготовка тез доповідей) – Метод кейс-стаді (аналіз реальних або змодельованих ситуацій, що дозволяє студентам застосовувати набуті знання у вирішенні практичних завдань) - Метод проектів (виконання студентами групових або індивідуальних проектів на основі професійних потреб)
РН-13. Володіє основами професійної культури, має здатність створювати й редагувати тексти професійного змісту державною та іноземною мовою.	– Вербальні методи (пояснення, розповідь, бесіда) – Наочні методи (презентація, демонстрація відеоматеріалів) – Практичні методи (виконання різних видів вправ, практичних завдань, творчі завдання) – Дискусійний метод (практикувати мову в ситуаціях, близьких до реальних професійних ситуацій) – Метод активного навчання (проведення ділових ігор, мозковий штурм, командна робота) – Методи самостійної роботи (анотування опрацьованого матеріалу, написання есе, підготовка тез доповідей) – Метод кейс-стаді (аналіз реальних або змодельованих ситуацій, що дозволяє студентам застосовувати набуті знання у вирішенні практичних завдань) - Метод проектів (виконання студентами групових або індивідуальних проектів на основі професійних потреб)

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/18

9. Методи контролю

Перевірка досягнення програмних результатів навчання здійснюється з використанням наступних методів.

Результат навчання	Методи контролю
РН-2. Обізнаний із елементами теоретичного й експериментального (пробного) дослідження в професійній сфері та методами їхньої реалізації.	– Усне опитування, участь у дискусії, відповіді на проблемні запитання – Перевірка виконання домашніх завдань, практичних завдань та вправ – Експрес-тестування – Перевірка виконання та захист індивідуальних завдань – Самооцінювання та взаємооцінювання – Перевірка виконання завдань модульного контролю – Залік
РН-9. Володіє різними видами аналізу художнього твору, визначає його жанрово-стильову своєрідність, місце в літературному процесі, традиції й новаторство, зв'язок твору із фольклором, міфологією, релігією, філософією, значення для національної та світової культури.	– Усне опитування, участь у дискусії, відповіді на проблемні запитання – Перевірка виконання домашніх завдань, практичних завдань та вправ – Експрес-тестування – Перевірка виконання та захист індивідуальних завдань – Самооцінювання та взаємооцінювання – Перевірка виконання завдань модульного контролю – Залік
РН-11. Порівнює мовні та літературні факти, явища, визначає їхні подібності й відмінності.	– Усне опитування, участь у дискусії, відповіді на проблемні запитання – Перевірка виконання домашніх завдань, практичних завдань та вправ – Експрес-тестування – Перевірка виконання та захист індивідуальних завдань – Самооцінювання та взаємооцінювання – Перевірка виконання завдань модульного контролю – Залік
РН-13. Володіє основами професійної культури, має здатність створювати й редагувати тексти	– Усне опитування, участь у дискусії, відповіді на проблемні запитання

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 19

Результат навчання	Методи контролю
професійного змісту державною та іноземною мовою.	– Перевірка виконання домашніх завдань, практичних завдань та вправ – Експрес-тестування – Перевірка виконання та захист індивідуальних завдань – Самооцінювання та взаємооцінювання – Перевірка виконання завдань модульного контролю – Залік

10. Оцінювання результатів навчання здобувачів вищої освіти

Оцінювання результатів навчання здобувачів вищої освіти з навчальної дисципліни здійснюється відповідно до Положення про оцінювання результатів навчання здобувачів вищої освіти у Державному університеті «Житомирська політехніка» та розподілу балів, що наведений нижче.

Система оцінювання результатів навчання здобувачів вищої освіти з навчальної дисципліни включає поточний та підсумковий контроль.

Поточний контроль проводиться для оцінювання рівня засвоєння знань, формування умінь і навичок здобувачів вищої освіти впродовж вивчення ними матеріалу модуля (змістових модулів) навчальної дисципліни. Поточний контроль здійснюється під час проведення навчальних занять.

Підсумковий контроль проводиться для підсумкового оцінювання результатів навчання здобувачів вищої освіти з навчальної дисципліни. Підсумковий контроль здійснюється після завершення вивчення навчальної дисципліни. Підсумковий контроль проводиться у формі заліку. Процедура складання заліку визначена у Положенні про організацію освітнього процесу у Державному університеті «Житомирська політехніка».

Розподіл балів з навчальної дисципліни

Види робіт здобувача вищої освіти	Кількість балів за семестр	
	денна форма	заочна форма
Виконання завдань поточного контролю	100	-
Підсумкова семестрова оцінка	100	-

Розподіл балів за виконання завдань поточного контролю

Види робіт здобувача вищої освіти	Кількість балів за семестр	
	денна форма	заочна форма
Виконання завдань під час навчальних занять	80	—

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/20

Види робіт здобувача вищої освіти	Кількість балів за семестр	
	денна форма	заочна форма
Виконання та захист індивідуальних самостійних завдань	20	—
Виконання науково-дослідної роботи та інших видів робіт (додаткові – заохочувальні бали): 1. Участь у студентських предметних олімпіадах, Всеукраїнському конкурсі студентських наукових робіт, грантах, науково-дослідних проектах 2. Підготовка наукових статей	до 10 до 10	—
Разом за виконання завдань поточного контролю	100	—

Розподіл балів за виконання завдань під час навчальних занять

Види робіт здобувача вищої освіти ¹	Кількість балів за семестр	
	денна форма	заочна форма
Відповіді (виступи) на заняттях	20	—
Участь у дискусії	10	—
Виконання поточних тестових завдань	30	—
Виконання та захист завдань, кейсів	20	—
Разом за виконання завдань під час навчальних занять	80	—

З метою застосування цілих чисел для оцінювання результатів роботи здобувачів під час навчальних занять протягом семестру може використовуватися 100-бальна шкала оцінювання щодо кожного окремо виду робіт. Розрахунок загальної кількості балів, які здобувач може набрати за результатами роботи під час навчальних занять за семестр, проводиться за формулою:

$$P_{\text{НЗ}} = \sum (P_i \times BK_i) \times K_{\text{НЗ}}, \quad (1)$$

де $P_{\text{НЗ}}$ – загальна кількість балів, набраних здобувачем за виконання завдань під час навчальних занять за семестр;

P_i – кількість набраних здобувачем балів за семестр за виконання i -го виду робіт під час навчальних занять (за 100-бальною шкалою);

BK_i – ваговий коефіцієнт за виконання i -го виду робіт під час навчальних занять. Значення вагових коефіцієнтів розраховуються шляхом ділення кількості балів, яка передбачена за виконання окремого виду робіт під час навчальних занять, на сумарну кількість балів за виконання усіх видів робіт під час навчальних занять за семестр;

$K_{\text{НЗ}}$ – коригувальний коефіцієнт, який визначається шляхом ділення кількості балів, що передбачена за виконання завдань під час навчальних занять за семестр, на 100 балів.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 21

На залік виносяться ключові питання за темами навчальної дисципліни, які вивчалися у відповідному семестрі.

Здобувач вищої освіти допускається до процедури підсумкового контролю у формі заліку, якщо протягом семестру за виконання завдань поточного контролю набрав 50 балів або більше.

Визнання результатів навчання, набутих у неформальній та/або інформальній освіті

Визнання результатів навчання, набутих у неформальній та/або інформальній освіті в рамках окремих тем навчальної дисципліни, здійснюється викладачем за зверненням здобувача вищої освіти та представленням документів, які підтверджують результати навчання (сертифікати, свідоцтва, скріншоти тощо). Рішення про визнання та оцінка за відповідну частину освітнього компонента приймається викладачем за результатами співбесіди зі здобувачем вищої освіти.

Визнання результатів навчання, набутих у неформальній та/або інформальній освіті в рамках цілого освітнього компонента, здійснюється за процедурою, яка визначена у Положенні про організацію освітнього процесу у Державному університеті «Житомирська політехніка».

Шкала оцінювання

Шкала ЄКТС	Національна шкала	100-бальна шкала
	Залік	
A	Зараховано	90-100
B	Зараховано	82-89
C		74-81
D	Зараховано	64-73
E		60-63
FX	Не зараховано	35-59
F		0-34

11. Глосарій

№ з/п	Термін державною мовою	Відповідник англійською мовою
1	Text Mining	текстовий майнінг
2	Natural Language Processing (NLP)	обробка природної мови
3	Tokenization	токенізація

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/ 22

№ з/п	Термін державною мовою	Відповідник англійською мовою
4	Stemming	стемінг
5	Lemmatization	лемматизація
6	Corpus	корпус текстів
7	Stop Words	стоп-слова
8	Part-of-Speech Tagging (POS Tagging)	визначення частин мови
9	Named Entity Recognition (NER)	розпізнавання іменованих сутностей
10	Feature Extraction	виділення ознак
11	Vectorization	векторизація
12	Bag of Words (BoW)	мішок слів
13	TF-IDF (Term Frequency-Inverse Document Frequency)	частота термів, зважена оберненою частотою документів
14	Word Embedding	векторне представлення слів
15	Word2Vec	модель Word2Vec
16	FastText	модель FastText
17	Latent Dirichlet Allocation (LDA)	латентне розміщення Діріхле
18	Topic Modeling	тематичне моделювання
19	Text Classification	класифікація тексту
20	Sentiment Analysis	аналіз тональності
21	Supervised Learning	навчання з учителем
22	Unsupervised Learning	навчання без учителя
23	Clustering	кластеризація
24	Cosine Similarity	косинусна схожість
25	Semantic Similarity	семантична схожість
26	Named Entity	іменована сутність
27	Phrase Extraction	виділення фраз
28	Bigram	біграма
29	N-gram	n-грам
30	Text Preprocessing	попередня обробка тексту
31	Noise Removal	видалення шуму

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/23

№ з/п	Термін державною мовою	Відповідник англійською мовою
32	Data Cleaning	очищення даних
33	Stopword Removal	видалення стоп-слів
34	Text Normalization	нормалізація тексту
35	Language Model	мовна модель
36	Sequence-to-Sequence Model (Seq2Seq)	модель послідовність-послідовність
37	Transformer	трансформер
38	BERT (Bidirectional Encoder Representations from Transformers)	двонаправлені уявлення від трансформерів
39	GPT (Generative Pre-trained Transformer)	генеративний попередньо натренований трансформер
40	Overfitting	перенавчання
41	Underfitting	недонавчання
42	Evaluation Metrics	метрики оцінювання
43	Precision	точність
44	Recall	повнота
45	F1-Score	метрика F1
46	Cross-validation	перехресна валідація
47	Train-Test Split	поділ на навчальну і тестову вибірки
48	Language Understanding	розуміння мови
49	Chatbot	чат-бот
50	Ethical Considerations	етичні аспекти

12. Рекомендована література

Основна література

1. Thakur K., Kumar V. An Overview of Text Mining: Application and Free Software Tools. Library Waves, 2020, no. 2(6), pp. 53-59.

2. Карпов І., Антоненко С. Огляд методів інтелектуального аналізу тексту. Актуальні проблеми автоматизації та інформаційних технологій. 2020. № 24. С. 40–46.

3. Муковський І.Т., Міщенко А.Г., Шевченко М.М. Інформаційно-аналітична діяльність у міжнародних відносинах: Навч. посіб. Київ : Кондор, 2021. 224 с

Допоміжна література

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/24

1. Bober, N., Kapranov, Y., Kukarina, A., Tron, T., & Nasalevych, T. (2021). British National Corpus in English language teaching of university students. *International Journal of Learning, Teaching and Educational Research*, 20(6), 174-193. <https://doi.org/10.26803/ijlter.20.6.9> (Scopus, Q3).

2. Chen A.P., Zhang C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 2014, vol. 275, pp. 314-347, 2014. doi: 10.1016/j.ins.2014.01.015.

3. Henriksson A., Moen H., Skeppstedt M., Daudaravicius V., Duneld M. Synonym extraction and abbreviation expansion with ensembles of semantic spaces. *Journal of biomedical semantics*, 2014, no. 1(5), p. 1, doi: 10.1186/2041-1480-5-6.

4. Joby P. J., Korra J. Accessing accurate documents by mining auxiliary document information. *Advances in Computing and Communication Engineering (ICACCE)*, Second International Conference on. IEEE, 2015, pp. 634-638. doi: 10.1109/ICACCE.2015.37.

5. Kapranov, Y. (2022). Antconc Corpus Manager and Its Possibilities for Keywords with Resilience Semantics Search. In R. Vasko (ed.) *Language. Culture. Discourse*. Kyiv. *Nat. ling. Un-t.* 22-32. <https://doi.org/10.15587/978-617-7319-60-2.2> (Scopus).

6. Kapranov, Y. (2022). Logical-Philosophical Reinterpretation of Retrospective Nature of AI Technology in the Modern Globalized World. *Logos*, 113, 35- 45. <https://doi.org/10.24101/logos.2022.71> (Scopus, Web of Science).

7. Kapranov, Y., & Tron, T. (2022). *Discourse Markers in Everyday English Communication: Translation and Corpus Verification: MANUAL for Higher Education Institutions that Carry Out Educational Activities at the First (Bachelor) and the Second (Master) Levels of Higher Education in the Specialty 035 Philology*. KNLU Publishing House.

8. Kapranov, Y., Cherkhava, O., Gromova, V., Reshetnyk, O. (2021). Methodological procedure of diagnosing behavioural stereotypes resilience of different language cultures representatives. *Amazonia Investiga*, 10(39). P. 186-193. <https://doi.org/10.34069/AI/2021.39.03.18> (Web of Science).

9. Kapranov, Y., Korolyova, A., Tolcheyeva, T., & Vasko, R. (2020). Human Language as a Natural Artifact of Planetary-Noospheric Mind: Coevolutionary-Macromutational Reinterpretation. *Amazonia Investiga*, 9(34), 17-23. <https://doi.org/10.34069/AI/2020.34.10.2> (Web of Science).

10. Kapranov, Y., Liashko, O., Cherkhava, O., Nasalevych, T., Riabukha, T. (2021). Methodological Algorithm for Interpreting Intertextuality within the Orthodox Sermon (Case Study of the English Orthodox Sermon Fragment). *Amazonia Investiga*, 10(39), 186-193. <https://doi.org/10.34069/AI/2021.43.07.12> (Web of Science).

11. Laxman A., Sujatha D. Improved method for pattern discovery in text mining. *International Journal of Research in Engineering and Technology*, 2013, no. 1(2), pp. 2321-2328. doi:10.15623/IJRET.2013.0210090.

12. Liao S. H., Chu P. H., Hsiao P. Y. Data mining techniques and applications - a decade review from 2000 to 2011. *Expert Systems with Applications*, 2012, no. 39(12),

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/25

pp. 11303-11311. doi:10.1016/j.eswa.2012.02.063.
https://doi.org/10.32782/mathematical-modelling/2022-5-2-5

13. Liashko, O., Bober, N., Kapranov, Y., Cherkhava, O., & Meleshkevych, L. (2022). Interpretation of Keywords as Indicators of Intertextuality in English New Testament Texts (Antconc Corpus Manager Toolkit). *Wisdom*, 22(2), 193- 207. https://doi.org/10.2423 4/wisdom.v22i2.736 (Scopus, Q2).

14. Rajendra R., Saransh V. A Novel Modied Apriori Approach for Web Document Clustering. *International Journal of Computer Applications*, 2013, vol. 33, pp. 159-171. doi:10.1007/978-81-322-2202-6_14.

15. Ramzan Talib, Muhammad Kashif Hanif, Shaeela Ayesha and Fakeeha Fatima. Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications*, no.11(7), 2016. doi: 10.14569/IJACSA.2016.071153.

16. Sheela S. K., Bharathi T. Analyzing Different Approaches of Text Mining Techniques and Applications. *International Journal of Computer Science Trends and Technology*, 2018, no. 4(6).

17. Sumathy K. L., Chidambaram M. Text mining: Concepts, applications, tools and issues – an overview. *International Journal of Computer Applications*, 2013, no. 4 (80), pp. 29-32. doi: 10.5120/13851-1685.

18. Vasko, R., Kapranov, Y., Korolyova, A., & Hryshchuk, Y. (2021). Transfer of Mathematical Formulas and Computer Algorithms into Macrocomparative Studies. 11th International Conference on Advanced Computer Information Technologies (ACIT): Proceedings, 642–647. (Scopus).

19. Wen Z., Yoshida T., Tang X. A study with multi-word feature with text classication. *Proceedings of the 51st Annual Meeting of the ISSS-2007*, Tokyo, Japan, 2007, vol. 51, p. 45.

20. Zhong N., Li Y., Wu S.T. Effective pattern discovery for text mining. *IEEE transactions on knowledge and data engineering*, 2012, no. 1(24), pp. 30-44. doi: 10.1109/TKDE.2010.211.

21. Капранов Я. В. “Діахронічна інтерпретація і верифікація ступенів споріднення ностратичних мов (на матеріалі етимонів *HuḲa, *wol[a], *wete, *phañ- (~ *phəñ-), *marV)” на здобуття наукового ступеня доктора філологічних наук зі спеціальності 10.02.17 – порівняльноісторичне і типологічне мовознавство. Захист відбувся на засіданні спеціалізованої вченої ради Д 26.054.02 у Київському національному лінгвістичному університеті 24 березня 2021 р. (Науковий консультант – д.ф.н., проф. Корольова А. В.).

22. Капранов, Я. & Тронь, Т. (2023). Основи лексикографічної теорії і практики: підручник для закладів вищої освіти, які здійснюють професійну підготовку майбутніх здобувачів першого (бакалаврського) рівня вищої освіти зі спеціальності 035 Філологія.

23. Капранов, Я. & Черхава, О. (2023). Основи лінгвістичної компаративістики та макрокомпаративістики: електронний підручник для закладів вищої освіти, які

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015			Ф-31.06- 05.01/014.021.1/Б/ОК26- 1-2024
	Випуск 1	Зміни 0	Екземпляр № 1	Арк 26/26

здійснюють професійну підготовку здобувачів другого (магістерського) рівня вищої освіти зі спеціальності 035 Філологія. Видавничий центр КНЛУ.

13. Інформаційні ресурси в Інтернеті

1. Засоби та методи аналізу неструктурованих даних [Електронний ресурс] – Режим доступу: <http://dspace.nbuv.gov.ua/bitstream/handle/123456789/150922/06-Rogushina.pdf?sequence=1> -
2. Контент аналіз [Електронний ресурс] – Режим доступу: <https://core.ac.uk/download/pdf/149247424.pdf>
3. Інтелектуальний аналіз тексту [Електронний ресурс] – Режим доступу: <https://actualproblems.dp.ua/article> › article ›
4. Alex Honchar. Sentiment analysis: solutions and applications survey [Електронний ресурс] – Режим доступу: <https://hp-analytics.medium.com/sentiment-analysis-solutions-and-applicationssurvey-9e52d3ea2ac7>
5. Shivam Bansal. A Comprehensive Guide to Understand and Implement Text Classification in Python [Електронний ресурс] – Режим доступу: <https://www.analyticsvidhya.com/blog/2018/04/acomprensive-guide-to-understand-and-implement-text-classification-in-python/>
6. Olivier Habimana. Sentiment analysis using deep learning approaches: an overview [Електронний ресурс] – Режим доступу: https://www.researchgate.net/publication/338351280_Sentiment_analysis_using_deep_learning
7. Paweł Cichosz. Naive Bayes classifier [Електронний ресурс] – Режим доступу: https://www.researchgate.net/publication/313991281_Naive_Bayes_classifier