

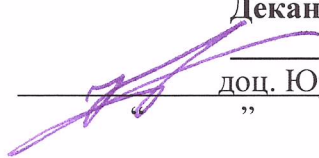
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА

Кафедра (циклова комісія) оптоелектроніки та інформаційних технологій

“ЗАТВЕРДЖУЮ”

Декан факультету

  
доц. Юрій ФУРГАЛА

”

2022 року

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Комп’ютерна лінгвістика та обробка природної мови

(шифр і назва навчальної дисципліни)

спеціальність 121 – Інженерія програмного забезпечення

(шифр і назва спеціальності)

спеціалізація \_\_\_\_\_

(назва спеціалізації)

факультет електроніки та комп’ютерних технологій

(назва інституту, факультету, відділення)

2022 – 2023 навчальний рік

Робоча програма \_\_\_\_\_ “Комп’ютерна лінгвістика та обробка природної мови”  
\_\_\_\_\_ для студентів

(назва навчальної дисципліни)

галузі знань \_\_\_\_\_ “12 – Інформаційні технології”  
за спеціальністю \_\_\_\_\_ “121 Інженерія програмного забезпечення”

Розробники: \_\_\_\_\_ проф., докт. фіз.-мат. наук Кушнір О. С.  
(вказати авторів, їхні посади, наукові ступені та вчені звання)

Робочу програму схвалено на засіданні кафедри (циклової комісії) \_\_\_\_\_  
\_\_\_\_\_ оптоелектроніки та інформаційних технологій

Протокол від “ 29 ” \_\_\_\_\_ 08 \_\_\_\_\_ 2022 року № \_\_\_\_\_

Завідувач кафедри \_\_\_\_\_ оптоелектроніки та інформаційних технологій



(Олег КУШНІР)

(підпис)

(прізвище та ініціали)

Ухвалено Вченою радою \_\_\_\_\_ факультету електроніки та комп’ютерних технологій

Протокол від “ 31 ” \_\_\_\_\_ 08 \_\_\_\_\_ 2022 року № 28/22

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА**

---

Кафедра (циклова комісія) оптоелектроніки та інформаційних технологій

**“ЗАТВЕРДЖУЮ”**

**Декан факультету**

доц. Юрій ФУРГАЛА

“ ” 2022 року

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Комп’ютерна лінгвістика та обробка природної мови

(шифр і назва навчальної дисципліни)

спеціальність 121 – Інженерія програмного забезпечення

(шифр і назва спеціальності)

спеціалізація \_\_\_\_\_

(назва спеціалізації)

факультет електроніки та комп’ютерних технологій

(назва інституту, факультету, відділення)

2022 – 2023 навчальний рік

Робоча програма \_\_\_\_\_ “Комп’ютерна лінгвістика та обробка природної мови”  
\_\_\_\_\_ для студентів

(назва навчальної дисципліни)

галузі знань \_\_\_\_\_ “12 – Інформаційні технології”

за спеціальністю \_\_\_\_\_ “121 Інженерія програмного забезпечення”

Розробники: \_\_\_\_\_ проф., докт. фіз.-мат. наук Кушнір О. С.

(вказати авторів, їхні посади, наукові ступені та вчені звання)

Робочу програму схвалено на засіданні кафедри (циклової комісії) \_\_\_\_\_

\_\_\_\_\_ оптоелектроніки та інформаційних технологій

Протокол від “\_\_\_\_\_” \_\_\_\_\_ 2022 року № \_\_\_\_\_

Завідувач кафедри \_\_\_\_\_ оптоелектроніки та інформаційних технологій

\_\_\_\_\_ (Олег КУШНІР)  
(підпис) (прізвище та ініціали)

Ухвалено Вченою радою \_\_\_\_\_ факультету електроніки та комп’ютерних технологій

Протокол від “\_\_\_\_\_” \_\_\_\_\_ 2022 року № \_\_\_\_\_

## 1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни
		<i>денна форма навчання</i>
Кількість кредитів – 5,0	Галузь знань <b>12 – Інформаційні технології</b>	Вибіркова дисципліна
Модулів – 1	Спеціальність <b>121 – Інженерія програмного забезпечення</b>	<i>Рік підготовки:</i> <b>4-й</b>
Змістових модулів – 3		<i>Семестр</i> <b>8-й</b>
Індивідуальне науково-дослідне завдання		
Загальна кількість годин – 150		
Тижневих годин для денної форми навчання: <i>Аудиторних:</i> VIII семестр – 4 <i>Самостійної роботи студента:</i> VIII семестр – 5		
	Освітньо-кваліфікаційний рівень: <b>бакалавр</b>	<i>Лекції</i> <b>32 год.</b>
		<i>Практичні, семінарські</i> –
		<i>Лабораторні</i> <b>32 год.</b>
		<i>Самостійна робота</i> <b>86 год.</b>
		<i>Індивідуальні завдання</i> –
		<i>Вид контролю:</i> <b>залік</b>

### Примітка.

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

для денної форми навчання 1,14

для заочної форми навчання – немає

## 2. Мета та завдання навчальної дисципліни

**Мета:** вивчення дисципліни «Комп'ютерна лінгвістика та обробка природної мови» націлене на ознайомлення студентів з теоретичними основами, алгоритмами та методами комп'ютерної лінгвістики та опрацювання природної мови.

**Завдання:** формування в студентів практичних навичок, які би дали змогу використовувати засвоєні знання, а також обирати, обґрунтовувати та ефективно застосовувати алгоритми, методи і прикладні програми для опрацювання природної мови.

У результаті вивчення цього курсу студент повинен:

**знати:** основні методи комп'ютерної лінгвістики та опрацювання природної мови, основні визначення, теорії, моделі та алгоритми опрацювання природної мови і опису лінгвістичних систем, інформаційного пошуку та інтелектуального аналізу текстових даних;

**вміти:** аналізувати моделі для опрацювання природної мови, створювати програмні продукти та працювати з готовими програмними продуктами, застосовувати комп'ютерну техніку для вирішення лінгвістичних задач, розробляти та реалізувати відповідні алгоритми, писати прикладні програми та користуватися ними.

Після вивчення даного курсу студенти набудуть таких Загальних та фахових компетентностей і Програмних результатів навчання:

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК3. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК4. Здатність спілкуватися іноземною мовою як усно, так і письмово.

ЗК6. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ФК16. Здатність ідентифікувати, класифікувати та формулювати вимоги до програмного забезпечення.

ФК18. Здатність розробляти архітектури, модулі та компоненти програмних систем.

ФК19. Здатність формулювати та забезпечувати вимоги щодо якості програмного забезпечення у відповідності з вимогами замовника, технічним завданням та стандартами.

ФК30. Здатність використовувати для розробки програмного забезпечення перспективні технології, зокрема, системи штучного інтелекту, Fog/Edge-обчислення тощо.

ФК32. Здатність здійснювати розробку програмного забезпечення використовуючи різні парадигми програмування (в тому числі паралельне, об'єктно-орієнтоване, функціональне програмування тощо)

ПРН 1. Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідкові ресурси і знання з урахуванням сучасних досягнень науки і техніки.

ПРН5. Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізу та математичного моделювання для розробки програмного забезпечення.

ПРН6. Уміння вибирати та використовувати відповідну задачі методологію створення програмного забезпечення.

Для вивчення курсу студенти потребують базових знань з дисциплін «Вища математика», «Дискретна математика», «Основи програмування», «Теорія алгоритмів», «Методи та технології обчислень», «Прикладна статистика та ймовірнісні процеси», «Об'єктно-орієнтоване програмування», «Бази даних», «Алгоритми і структури даних».

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

#### **МОДУЛЬ 1**

#### **Змістовий модуль 1. Поняття, методи та продукти комп'ютерної лінгвістики**

##### **Тема 1. Вступ. Лінгвістика та її структура (2 год.).**

Передмова. Мета курсу та цільова аудиторія. Зв'язки комп'ютерної лінгвістики з галузями інформатики та систем штучного інтелекту.

Роль обробки природної мови. Лінгвістика та її структура. Поняття та межі галузі комп'ютерної лінгвістики. Поняття слова.

Використання підходів і методів фундаментальної науки. Сучасний стан прикладних досліджень з комп'ютерної лінгвістики.

### **Тема 2. Основні поняття лінгвістики (4 год.).**

Загальні поняття про мову. Мова і мислення. Мова як знакова система. Мова і мовлення. Мовна структура. Рівні мови.

Фонетика. Звуковий характер мови. Фонологія. Поняття про фонемі. Звуки і букви. Поняття про алфавіт.

Морфологія. Основні поняття. Морфема. Види морфем. Формальні моделі морфології. Морфологічні словники.

Синтаксис. Основні поняття. Словосполучення. Речення. Синтаксичні відносини. Керування. Узгодження. Примикання. Комунікативна організація речення.

Семантика. Основні поняття. Значення. Види значень. Семантичні відносини. Парадигматичні та синтагматичні відносини. Формальні моделі семантики.

Основи лексикографії. Семантичні мережі. Фрейми. Лексико-семантичні комплекси.

Прагматика. Основні поняття. Види прагматичних знань. Проблема розуміння текстів. Умови цільового розуміння текстів.

### **Тема 3. Розвиток ідей, теорій, підходів і методів комп'ютерної лінгвістики (2 год.)**

Структурний підхід в комп'ютерній лінгвістиці. Внесок Н. Хомського у комп'ютерну лінгвістику.

Проста контекстно-вільна граматики. Трансформаційні граматики.

Лінгвістичні дослідження після Н. Хомського. Валентності та тлумачення.

Труднощі теорії Н. Хомського.

Грамматика структури фраз, керована головною складовою. Ідея уніфікації.

Теорія ЗМІСТ ↔ ТЕКСТ: мова як багатоступеневий перетворювач.

Моделі керування. Дерева залежностей. Семантичні зв'язки.

### **Тема 4. Продукти комп'ютерної лінгвістики: досягнення та проблеми (4 год.)**

Класифікація прикладних лінгвістичних систем. Автоматична розстановка переносів. Перевірка орфографії. Перевірка граматики. Перевірка стилістики.

Пошук слів і словосполучень.

Інформаційні пошукові системи.

Автоматичний переклад.

Інтерфейс на природній мові.

Видобування фактичних даних з текстів. Генерація тексту.

Системи розуміння мови та пов'язані з ними системи.

## **Змістовий модуль 2. Основні теорії та моделі, покладені в основу алгоритмів комп'ютерної лінгвістики**

### **Тема 5. Мова як двонаправлений перетворювач змісту та тексту (2 год.).**

Можливі точки зору стосовно природної мови. Мова як двонаправлений перетворювач. Поняття тексту і змісту. Два способи представлення змісту.

Розкладання і атомізація змісту. Неоднозначність “ картування ” ЗМІСТ ↔ ТЕКСТ: синоніми і омоніми. Подальша інформація про омонімію.

Багатоетапність перетворення ЗМІСТ ↔ ТЕКСТ. Переклад як багатоступінчасте перетворення.

### **Тема 6. Лінгвістичні знаки (2 год.).**

Два боки поняття знака. Лінгвістичний знак.

Лінгвістичний знак у теорії ЗМІСТ ↔ ТЕКСТ.

Лінгвістичний знак у теорії граматики структури фраз, керованої головною складовою.

Змістовний знак: умовність чи дане природою?

Порівняння генеративних ідей, ідей з теорії ЗМІСТ ↔ ТЕКСТ та ідей обмежень.

### **Тема 7. Лінгвістичні моделі (4 год.).**

Поняття моделювання. Нейролінгвістичні моделі. Психолінгвістичні моделі.

Функціональні моделі мови. Дослідження лінгвістичних моделей.

Спільні риси сучасних моделей мови. Особливості моделі ЗМІСТ ↔ ТЕКСТ. Зредуковані моделі. Практична потреба в лінгвістичних моделях.

Аналогії в природних мовах. Емпіричні підходи проти раціоналістичних підходів.

Обмеженість сфери застосування сучасних лінгвістичних теорій.

### **Змістовий модуль 3. Аналіз, розпізнавання та синтез природної мови, машинний переклад та комп'ютерна лексикографія**

#### **Тема 8. Автоматичне введення звуків мови, аналіз та розпізнавання мови комп'ютером (4 год.).**

Акустична будова звуків мови. Загальні принципи організації мовної комунікації. Будова і принципи роботи мовнісного тракту людини. Теорія сприйняття звуків мови людиною.

Аналого-цифрове перетворення мовних сигналів.

Проблеми автоматичного розпізнавання усного мовлення.

Методи аналізу та розпізнавання усного мовлення. Первинна обробка мовних сигналів. Методи розпізнавання окремих звуків. Звуковий спектрограф.

Спектрально-часовий аналіз мовлення. Бінарна селекція звукових елементів мови.

Методика формантного аналізу мови.

Розпізнавання мовлення із залученням лінгвістичної інформації.

Проблема автоматичного розпізнавання усного мовлення.

#### **Тема 9. Синтез мови та мовні технології (4 год.).**

Розвиток систем синтезу усного мовлення. Метод кодування сигналів.

Фонетичний синтез мовлення за допомогою формантних синтезаторів.

Синтез мовлення за методом предикативного кодування.

Алгоритм синтезу усного мовлення за текстом.

Загальна структура синтезатора.

Лінгвістичний, просодичний, фонетичний, акустичний та компіляційний синтезатори.

Сучасні технології усного мовлення.

#### **Тема 10. Машинний переклад і комп'ютерна лексикографія (4 год.).**

Загальні положення машинного перекладу. Оцінка та критерії якості перекладу. Типові помилки машинного перекладу.

Аналіз теоретичних моделей. Структура систем перекладу.

Алгоритми аналізу та синтезу.

Види перекладу та характеристики систем перекладу. Сучасний стан машинного перекладу.

Предмет, методи та теорія комп'ютерної лексикографії. Проблеми комп'ютерної лексикографії.



#### 4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин					
	Денна форма					
	Усього	у тому числі				
		л.	пр.	лаб.	інд.	с. р.
1	2	3	4	5	6	7
<b>МОДУЛЬ 1</b>						
<b>Змістовий модуль 1. Поняття, методи та продукти комп'ютерної лінгвістики</b>						
Тема 1. Вступ. Лінгвістика та її структура	8	3		–		5
Тема 2. Основні поняття лінгвістики	8	3		–		5
Тема 3. Розвиток ідей, теорій, підходів і методів комп'ютерної лінгвістики	19	3		6		10
Тема 4. Продукти комп'ютерної лінгвістики: досягнення та проблеми	19	3		6		10
<i>Разом за змістовий модуль 1</i>	<b>54</b>	<b>12</b>		<b>12</b>		<b>30</b>
<b>Змістовий модуль 2. Основні теорії та моделі, покладені в основу алгоритмів комп'ютерної лінгвістики</b>						
Тема 5. Мова як двонаправлений перетворювач змісту та тексту	13	3		–		10
Тема 6. Лінгвістичні знаки	19	3		6		10
Тема 7. Лінгвістичні моделі	20	4		6		10
<i>Разом за змістовий модуль 2</i>	<b>52</b>	<b>10</b>		<b>12</b>		<b>30</b>
<b>Змістовий модуль 2. Аналіз, розпізнавання та синтез природної мови, машинний переклад та комп'ютерна лексикографія</b>						
Тема 8. Автоматичне введення звуків мови, аналіз та розпізнавання мови комп'ютером	18	4		4		10
Тема 9. Синтез мови та мовні технології	15	3		2		10
Тема 10. Машинний переклад і комп'ютерна лексикографія	13	3		4		6
<i>Разом за змістовий модуль 3</i>	<b>46</b>	<b>10</b>		<b>10</b>		<b>26</b>
<b>Усього годин</b>	<b>150</b>	<b>32</b>		<b>32</b>		<b>86</b>

#### 5. Теми семінарських занять

Семінарських занять за курсом не передбачено.

#### 6. Теми практичних занять

Практичних занять за курсом не передбачено.

## 7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
VIII семестр		
1.	Препроцесинг текстових документів	2
2.	Визначення семантичного навантаження тексту за параметрами кластеризації слів	2
3.	Опрацювання природних мов на різних лінгвістичних рівнях за допомогою засобів Python	2
4.	Фонетичний аналіз і силабіфікація текстів	3
5.	Програмування задач морфологічного синтезу дієслівних форм	3
6.	Рангові залежності та розподіли ймовірності для прізвищ та імен	2
7.	Мережеві методи визначення ключових слів у текстах	3
8.	Закони Ціпфа, Парето та Гіпса для слів у текстах	3
9.	Практичні рецепти бінування для словника корпусу текстів	2
10.	Статистика n-грам у текстах і корпусах текстів	2
11.	Вивчення рандомних моделей текстів	2
12.	Метод TF-IDF	2
13.	Кластеризаційні методи встановлення ключових слів у текстах	2
14.	Вирішення задач розпізнавання за допомогою пакету FineReader	2
	<b>Разом</b>	<b>32</b>

## 8. Самостійна робота

№ з/п	Назва теми	Кількість годин
VIII семестр		
1.	Основні наукові напрямки комп'ютерної лінгвістики та її зв'язки з іншими науками. Комп'ютерна лінгвістика і прикладні завдання мовознавства. Інформатика і комп'ютерна лінгвістика.	4
2.	Вибір системи машинного перекладу (СМП). Порівняння функціональних можливостей СМП.	2
3.	Порівняння інтерфейсів СМП. Можливості інтеграції систем. Ввід, вивід і обробка тексту.	2
4.	Налаштування словників. Заборона перекладу слів. Статистична обробка перекладу.	2
5.	Порівняння якості перекладу різними системами.	2
6.	Англо-український переклад. Українсько-англійський переклад.	2
7.	Оптимізація роботи СМП.	2
8.	Підготовка тексту для машинного перекладу. Введення тексту в комп'ютер. Орфографічна перевірка тексту.	2
9.	Загальний підхід до налаштування СМП. Відкриття вихідних документів. Підключення і конфігурація словників.	2
10.	Виявлення слів, які не перекладаються. Задавання варіантів перекладу слів. Переклад пробного фрагменту і остаточне налаштування СМП.	2
11.	Постредагування. Змістовна, нормативна і факультативна правка. Усування слів, які не перекладаються.	2

12.	Неправильний переклад слів. Помилки в перекладі зворотів. Порушення погодження слів. Нерозв'язана омонімія. Помилки в перекладі пасивних конструкцій.	2
13.	Помилки у порядку і сполучуваності слів. Помилки в перекладі дієслова to be. Неправильний переклад іменникових груп. Неправильний переклад безособових форм дієслова. Методичні помилки оригіналів.	2
14.	Загальні прийоми роботи з інтерфейсом СМП.	2
15.	Електронні словники і робота з ними. Система електронних словників Polyglossum II.	2
16.	Інсталяція електронних словників.	2
17.	Налаштування шрифтів. Запуск програми. Пошук слів у словниках. Перенос інформації до тесту.	2
18.	Макрокоманди.	2
19.	Системи електронних словників Lingvo. Налаштування інтерфейсу. Перегляд словникових статей. Редагування словникових статей.	2
20.	Stylus для Windows 98/NT. Редагування тексту в системі Stylus. Спільна робота Stylus та інших додатків Windows.	2
21.	PARS для Windows. Спільна робота PARS і MS Word.	2
22.	Socrat для Windows.	2
23.	PROMT Windows 98/NT.	2
24.	Машинний переклад: елементи хронології, елементів періодизації стратегій і елементи типології систем. Основні області використання машинного перекладу.	2
25.	Деякі відомі системи машинного перекладу. Перспективи розвитку систем машинного перекладу.	2
26.	Теорія і методика викладання мови. Граматико-перекладний метод у вивченні іноземної мови. Прямий метод вивчення іноземної мови.	2
27.	Аудіолінгвальний і аудіовізуальний методи вивчення іноземної мови. Комунікативно-орієнтований підхід у вивченні іноземної мови.	2
28.	Оптимізація “спілкування” з комп'ютером: системи обробки мов.	2
29.	Теорія і практика інформаційно-пошукових систем. Основні поняття інформаційного пошуку.	2
30.	Типи інформаційно-пошукових систем. Інформаційно-пошукові мови.	2
31.	Моделювання мовленнєвої діяльності. Основні інформаційні елементи створення тексту. Моделювання мовленнєвої діяльності в лінгвістиці.	2
32.	Інформаційні елементи тексту в усній та письмовій формі. Автоматичне індексування, реферування й анотування. Загальні принципи пошуку та одержання інформації в пошукових системах.	2
33.	Інформаційна оцінка знакових елементів аркуша. Обробка письмової лінгвістичної інформації на рівні словоформ, слів, словосполучень, речення, тексту.	2
34.	Машина морфологія. Методи автоматичного морфологічного аналізу тексту. Морфологічний аналіз зі словником словоформ.	2
35.	Омографія основ і афіксів. Морфологічний аналіз методом логічних множин.	2
36.	Автоматичний синтаксичний і семантичний аналіз тексту.	2
37.	Когнітивна наука як наукова дисципліна моделювання людського мислення. Штучний інтелект.	2
38.	Основні принципи створення інтелектуальних систем. Логічні закони в інтелектуальних системах. Евристика.	2
39.	Мова програмування PROLOG.	2
40.	Розпізнавання образів.	2

41.	Лінгвістичні підходи в експертних системах.	2
42.	Опис експертних систем аналізу навчання мови в школі.	2
	<b>Разом</b>	<b>86</b>

## 9. Індивідуальні завдання

Індивідуальних завдань за курсом не передбачено.

## 10. Методи навчання

Вивчення даної дисципліни передбачає лекційний курс і лабораторний практикум.

## 11. Методи контролю

Контроль засвоєння матеріалу включає поточний контроль (колоквіуми за трьома змістовими модулями – 45 балів) і оцінку відповідей та результатів роботи на лабораторних заняттях (55 балів), із підсумовуванням результатів на заліку. Сумарна оцінка за вивчення предмету виставляється за 100-бальною шкалою.

## 12. Розподіл балів, що присвоюються студентам

Поточне тестування та самостійна робота			Робота на лабораторних заняттях	Сума
Змістовий модуль 1	Змістовий модуль 2	Змістовий модуль 3		
14	15	15	56	100

### Розподіл балів, що присвоюються студентам на лабораторній роботі

№	Об'єкт та знання	Оцінка, бал	Примітки
1	Підготовка до виконання роботи	0,5	Оцінюється на поточному занятті
2	Виконання роботи	0,5	Оцінюється на поточному занятті
3	Звіт	2,0	Оцінюється на наступному занятті після виконання роботи
4	Захист звіту	1,0	Через місяць після виконання роботи максимальна можлива оцінка знижується до 2,0 бала
	Разом	4,0	

### Шкала оцінювання: університету, національна та ECTS

Оцінка в балах	Оцінка ECTS	Визначення	За національною шкалою	
			Екзаменаційна оцінка, оцінка з диференційованого заліку	Залік
90–100	A	Відмінно	Відмінно	Зараховано
81–89	B	Дуже добре	Добре	
71–80	C	Добре		
61–70	D	Задовільно	Задовільно	
51–60	E	Достатньо		

### **13. Методичне забезпечення**

Для засвоєння матеріалів навчальної дисципліни студентами наявне все необхідне методичне забезпечення, зокрема наочний матеріал в електронному форматі для проведення лекційних занять, база запитань і завдань для контролю рівня знань, а також методичні рекомендації для виконання лабораторних робіт.

### **14. Рекомендована література**

#### **Базова**

1. Волошин В. Г. Комп'ютерна лінгвістика / В. Г. Волошин. – Суми : Університетська книга, 2004. – 382 с.
2. Delmonte R. Computational Linguistic Text Processing / New York: Nova Science Publishers, 2009 – 382 p.
3. 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.
4. Bolshakov I. Computational linguistics. Models, resources, applications / I. Bolshakov, A. Gelbukh. – Mexico : Ciencia de la Computacion, 2004. – 198 p.
5. Clark A. The Handbook of Computational Linguistics and Natural Language Processing / A. Clark, C. Fox, S. Lappin. – Chichester: John Wiley and Sons, 2010. – 801 p.
6. Мирам Г. Алгоритмы перевода: Вступительный курс по формализации перевода / Г. Мирам. – Киев : Эльга, Ника-Центр, 2004. – 176 с.
7. Хархалис Р. И. Компьютерный перевод иностранных текстов / Р. И. Хархалис. Киев : Терези, 1998. – 193 с.

#### **Допоміжна**

8. Manning C. D. Foundations of statistical natural language processing / Manning C. D., Schutze H. – London: The MIT Press Cambridge, 1999. – 680 p.
9. Espitia D. Universal and non-universal text statistics: Clustering coefficient for language identification / D. Espitia, H. L. Ridaura // Physica A. – 2020. – Vol. 553. – 123905 (25 pp.).
10. Pilgrim C. Bias in Zipf's law estimators / C. Pilgrim, T. T. Hills // Scientific Reports. – 2021. – Vol. 11. – 17309 (12 p.).

### **15. Інформаційні ресурси**

1. <https://gutenberg.org/>
2. <https://github.com/>