



# КОМП'ЮТЕРНА ЛІНГВІСТИКА

## Силабус - Робоча програма навчальної дисципліни

### Реквізити навчальної дисципліни

Рівень вищої освіти	Перший (бакалаврський)
Галузь знань	12 Інформаційні технології
Спеціальність	121 Інженерія програмного забезпечення
Освітня програма	Інженерія програмного забезпечення інформаційних систем
Статус дисципліни	Вибіркові компоненти циклу професійної підготовки
Кількість кредитів	4
Рік підготовки, семестр	4 курс, весняний семестр
Форма навчання	Очна(денна)/ дистанційна
Розклад занять	Лекція – щотижня (36 годин) Лабораторне заняття – 1 раз на 2 тижні (18 годин)
Мова викладання	Українська
Керівник курсу / викладачі	Лектор: к.т.н. Фіногенов Олексій Дмитрович Лабораторні: Ромашкевич Яна Олександрівна
Контактна інформація	Кафедра інформатики та програмної інженерії (корп. 13, ауд. 23), e-mail: iri@kpi.ua Телефон:+380 44 204 90 45 Фіногенов Олексій Дмитрович e-mail: <a href="mailto:fenyatrashbox@gmail.com">fenyatrashbox@gmail.com</a> Telegram: @vse_F_AD Ромашкевич Яна Олександрівна e-mail: <a href="mailto:demoiselle.yana@gmail.com">demoiselle.yana@gmail.com</a> Telegram: @DemoiselleYana
Розміщення курсу	немає

### Програма навчальної дисципліни

#### 1. Опис навчальної дисципліни

Збільшення та урізноманітнення інформаційних потоків та мінливість новин, їх швидке «старіння», разом з розширенням меж використання аналітичної інформації в оперативному та довгостроковому горизонті прийняття рішень призводять до необхідності розуміння та принципів вирішення задач в межах лінгвістичних аспектів та її технічному аналогу – комп'ютерній лінгвістиці (КЛ). Актуальними питаннями КЛ є дослідження авторства текстів, аналіз емоційного забарвлення текстів, пошук та класифікація документів, генерація тестів/мовлення, автоматичний переклад та реферування, створення корпусів текстів для машинного навчання та багато інших задач. Метою навчальної дисципліни є посилення загальних та професійних компетенцій студентів:

- ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.
- ЗК06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- ФК07. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.
- ФК20. Здатність застосовувати існуючі методи та алгоритми для розв'язання практичних задач обробки даних та розпізнавання
- ФК21. Здатність обґрунтовувати вибір технічної структури та розробляти відповідне програмне забезпечення, що входить до складу технології інтернету речей

Після опанування курсу студенти здатні продемонструвати такі *результати навчання*:

- ПР01. Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки.
- ПР13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань
- ПР14. Застосовувати на практиці інструментальні програмні засоби доменного аналізу, проектування, тестування, візуалізації, вимірювань та документування програмного забезпечення.
- УМ02. Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки
- УМ12. Застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.

Особливістю дисципліни є її комплексний характер, націленість на формування у майбутніх інженерів навичок формалізації практичних задач, вибору методів рішення, врахування як технічних так і інших впливових факторів та аналізу отриманих результатів.

## **2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)**

Дисципліна базується на знаннях, отриманих студентами у процесі вивчення курсів «Алгоритми та структури даних», «Теорія ймовірностей», «Методи та системи прийняття рішень».

Отриманні результати навчання студенти зможуть застосовувати під час підготовки кваліфікаційної бакалаврської роботи та в подальшій фаховій діяльності.

## **3. Структура та зміст навчальної дисципліни**

Надається структура дисципліни (кількість освітніх компонентів, семестровий розподіл годин тощо) та перелік розділів і тем всієї **дисципліни**.

Семестр	Всього годин	Лекції	Лаб.	СРС	Контрольні заходи			
					МКР-1	Реф-0	КК-0	Залік
8	120	36	18	66	МКР-1	Реф-0	КК-0	Залік

СРС – самостійна робота студента; МКР – модульна контрольна робота; Реф – реферат; КК – календарний контроль

Навчальний матеріал курсу складається з 4 розділів:

Розділ 1. Вступ до комп'ютерної лінгвістики.

Розділ 2. Аналіз тексту.

Розділ 3. Обробка и перетворення тексту.

Розділ 4. Практичні задачі комп'ютерної лінгвістики.

### **Навчальний контент**

## **4. Логіка опанування навчальної дисципліни (освітнього компонента)**

Лекційний матеріал включає основні теоретичні дані та практичну складову задач комп'ютерної лінгвістики. Структура курсу побудована на поступовому переході від підбору та аналізу тексту до інших аспектів мови: перекладу, синтезу та розпізнавання, шифрування та кодування.

Лабораторні роботи включають опанування способів вирішення ряду практичних задач КЛ.

## **Розділ 1. Вступ до комп'ютерної лінгвістики.**

*Тема 1.1. Стан науки «комп'ютерна лінгвістика»*

**Лекція 1.** Організаційні питання. Історичні, соціокультурні, географічні аспекти.

**Лекція 2.** Корпусна лінгвістика.

## **Розділ 2. Аналіз тексту.**

*Тема 2.1. Авторство тексту.*

**Лекція 3.** Авторство тексту.

**Лекція 4.** Статистичні параметри тексту.

**Лабораторна робота 1.** Ідентифікація автора тексту за N-грамами.

*Тема 2.2. Канонічний вигляд тексту.*

**Лекція 5.** Канонічний вигляд тексту.

**Лекція 6.** Словники.

**Лекція 7.** Тональність тексту.

**Лабораторна робота 2.** Визначення тональності тексту.

**Лекція 8.** Латентно-семантичний аналіз.

**Лекція 9.** Модульна контрольна робота.

*Тема 2.3. Подібність документів.*

**Лекція 10.** Подібність документів. Плагіат.

**Лабораторна робота 3.** Визначення близькості документів.

## **Розділ 3. Обробка і перетворення тексту.**

*Тема 3.1. Обробка тексту.*

**Лекція 11.** Семантичні онтології.

**Лекція 12.** Списки Сводеша.

*Тема 3.2. Перетворення тексту.*

**Лекція 13.** Переклад тексту.

**Лекція 14.** Шифрування та кодування.

## **Розділ 4. Практичні задачі комп'ютерної лінгвістики.**

*Тема 4.1. Задачі, що пов'язані з текстом.*

**Лекція 15.** Генерація тексту.

**Лекція 16.** Розпізнавання тексту.

**Лекція 17.** Розпізнавання та синтез мовлення.

**Лекція 18.** Приклади досліджень з курсу «комп'ютерна лінгвістика».

## 5. Самостійна робота студента

Самостійна робота студента (66 годин) передбачає підготовку до аудиторних занять та контрольних заходів, проведення розрахунків за даними, отриманими на лабораторних роботах.

Розподіл годин СРС: підготовка до лекції – 1 година; підготовка до виконання та захисту лабораторної роботи – 5 годин; підготовка до МКР – 3 години.

### Політика та контроль

## 6. Політика навчальної дисципліни (освітнього компонента)

6.1 Відвідування лекційних, лабораторних занять:

- відвідування лекційних занять передбачається у синхронно/асинхронному режимі;
- з огляду на спрямованість ЛР для виконання на ЕОМ, відвідування лабораторних занять вимагає присутність на постановочній (теоретичній) частині лабораторної роботи або консультацій, та під час захисту.

6.2 Правила поведінки на заняттях.

6.2.1 В дистанційному (online) режимі:

- посилання на заняття в Zoom надаються в групі Telegram (що попередньо створена, старостам груп надані права на включення до групи студентів) перед проведенням заняття;
- під час лекції відео- та аудіо-канали студентів повинні бути відключені. Дозволяється вмикати канали під час звернення до викладача за умови реакції «піднята рука» щодо пояснень матеріалу під час лекції, загальних запитань викладача до аудиторії, відповіді на питання, які були до конкретного студента.

6.2.2 В аудиторному (offline) режимі:

- мобільні телефони повинні бути відключені або поставлені на безшумний режим;
- питання «з місця» задаються за умови піднятої руки та дозволу викладача.

6.3 Правила захисту лабораторних робіт:

- перед захистом лабораторні роботи надаються в електронному вигляді на пошту вказану викладачем з темою листа за шаблоном. Файл лабораторної роботи надається в форматах \*.doc, \*.docx або \*.pdf – методом вкладання файлу в лист. Посилання на гугл-диск або інші ресурси не дозволяються.
- захист лабораторної роботи складається з двох етапів: перевірки оформлення звіту та відповідей на запитання викладача;
- викладач перевіряє варіант завдання, обов'язкові елементи оформлення звіту лабораторної роботи, результати дослідження, графічні елементи (якщо передбачено), висновки.
- теоретична підготовка перевіряється у вигляді запитань викладача щодо теоретичних тверджень. Що використовуються у лабораторній роботі, цілей дослідження, вимоги пояснень щодо особливостей проведення експерименту, відмінностей при зміні початкових даних тощо.
- за результатами захисту викладач виставляє комплексну оцінку звіту та теоретичної підготовки студента.

6.4 Політика дедлайнів та перескладань

- дедлайни здачі лабораторних робіт та модульної контрольної роботи доводяться до студентів під час першого заняття та передбачають зменшення балів за несвоєчасну здачу;
- лабораторні роботи не перескладаються;

– модульна контрольна робота може бути перескладена не пізніше ніж за 3 тижні до закінчення навчального семестру за розкладом, що визначений викладачем, з врахуванням штрафних балів за несвоєчасне виконання.

6.5 Політика щодо академічної доброчесності:

– при виявленні плагіату (привласнення результатів чужої роботи) в звітах лабораторних робіт – оцінка за звіт автоматично визначається як рівна 0.

– при виявленні плагіату в модульній контрольній роботі – оцінка автоматично визначається як рівна 0.

## 7. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

**Поточний контроль:** захист лабораторних робіт, МКР.

**Календарний контроль:** не проводиться.

**Семестровий контроль:** залік.

Умови допуску до семестрового контролю: зарахування усіх лабораторних робіт та МКР.

**Остаточне оцінювання:**

Кількість балів	Оцінка
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

### 7.1 Розподіл балів за видами робіт

Сума вагових балів контрольних заходів протягом семестру складає (стартовий рейтинг):

$$R_C = 25 (\text{МКР}) + 75 (3 \text{ ЛР} * 25) = 100 \text{ балів.}$$

Таким чином, рейтингова шкала з дисципліни складає  $R = R_C = 100$  балів.

### 7.2 Критерії нарахування балів за результатами роботи студента:

7.2.1 Бали за виконання лабораторних робіт:

– за умови правильних результатів і своєчасного представлення роботи «25» балів;

– за умови невиконання (зниження) показника хоча б з однієї позиції «– 3» бали.

Максимальна кількість балів за всі лабораторні роботи дорівнює 75 балам.

Лабораторна робота	1	2	3
Виконання	15	15	15
Захист	10	10	10
<b>Загальний бал</b>	<b>25</b>	<b>25</b>	<b>25</b>
Термін виконання (тиждень)	1-6	4-6	7-9

7.2.2 Модульна контрольна робота

Критерії оцінювання МКР:

– “відмінно”, повна (вірна) відповідь (не менше 90% потрібної інформації) – 22-25 балів;

– “добре”, достатньо повна відповідь (не менше 75% потрібної інформації), або повна (вірна) відповідь з незначними помилками – 17-21 балів;

– “задовільно”, неповна відповідь (не менше 60% потрібної інформації) та незначні помилки – 13-16 балів;

- “незадовільно”, незадовільна відповідь (неправильний розв’язок задачі), потребує обов’язкового повторного написання 0-12 балів;

Відсутність на МКР без поважної причини та/або при повторному написанні: «-5» балів.

## 8. Додаткова інформація з дисципліни

### Базова література:

1. Карпіловська Є.А. Вступ до прикладної лінгвістики: комп’ютерна лінгвістика. Підручник. – Донецьк: ТОВ "Юго-Восток, Лтд", 2006. – 188 с.
2. Волошин В.Г. Комп’ютерна лінгвістика: Навчальний посібник. – Суми: ВТД "Університетська книга", 2004. – 382 с.
3. Карпіловська Є.А., Ситар Г.В. Теоретичні та прикладні проблеми комп’ютерної лінгвістики. Навчально-методичний посібник з курсу "Вступ до комп’ютерної лінгвістики" для студентів спеціальності 7.030.505 "Прикладна лінгвістика" / За ред. А.П.Загнітка. – Донецьк: ДонНУ, 2006. – 73 с.

### Додаткова література:

4. Кузьменко Д.М. Комп’ютерна лінгвістика і Web 2.0. Київ : Наукова думка, 2009. 250 с.
5. Міщенко А.Л. Сучасні методи, напрямки й здобутки комп’ютерної лінгвістики. Наукові записки Кіровоградського державного педагогічного університету імені В. Винниченка. Вип. 95(2). Серія «Філологічні науки», 208 (мовознавство). Кіровоград : РВВ КДПУ ім. В. Винниченка, 2011. С. 122–233.
6. Cecile L. Paris, William R. Swartout, William C. Mann Natural Language Generation in Artificial Intelligence and Computational Linguistics. Springer Science & Business Media, 14.03.2013. 404 p.
7. Biskub I. Applied and Computational Linguistics : підручник (англ. мовою) / I. Biskub. – Луцьк : РВВ «Вежа» Волин. держ. ун-ту ім. Лесі Українки, 2007. – 304 с.
8. 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.).

### Робочу програму навчальної дисципліни (силабус):

Склав доцент кафедри інформатики та програмної інженерії ФІОТ, к.т.н. **О.Д. Фіногенов**

Ухвалено на засіданні кафедри ІПІ ФІОТ, протокол № 16 від 29.05.2024р.

Погоджено на засіданні методичної комісії ФІОТ, протокол № 10 від 21.06.2024 р.