

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ
КАФЕДРА МАТЕМАТИЧНОЇ ІНФОРМАТИКИ**

«ЗАТВЕРДЖУЮ»
Заступник декана
з навчальної роботи
Олена КАШПУР
« 12 » лютого 2021 року

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
ОСНОВИ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ.**

для студентів

| | |
|--|--|
| галузь знань | 12 «Інформаційні технології» |
| спеціальність | 122 «Комп'ютерні науки» |
| освітній рівень | бакалавр |
| освітня програма | «Комп'ютерні науки» |
| вид дисципліни | вибіркова |
| вибірковий блок | «Інтелектуальні інформаційні технології» |
| Форма навчання | денна |
| Навчальний рік | 2023/2024 |
| Семестр | 7 |
| Кількість кредитів ECTS | 4 |
| Мова викладання, навчання та оцінювання | українська |
| Форма заключного контролю | іспит |

Викладачі: к.ф.-м.н., асистент Тарануха В.Ю. (лекції)

Пролонговано: на 20__/20__ н.р. _____ (_____) «__» 20__ р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____ (_____) «__» 20__ р.
(підпис, ПІБ, дата)

КИЇВ – 2021

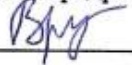
© Тарануха В.Ю., 2021

Розробник: Тарануха Володимир Юрійович, к.ф.-м.н., асистент кафедри «Математичної Інформатики»

Терещенко Василь Миколайович, д.ф.-м.н., професор, завідувач кафедри «Математичної Інформатики»

ЗАТВЕРДЖЕНО

Завідувач кафедри математичної інформатики

 Василь ТЕРЕЩЕНКО

Протокол № 6 від «11» 02 2021 р.

Схвалено Гарантом освітньо-професійної програми «Інформатика»

«11» лютого 2021р.  Людмила ОМЕЛЬЧУК

Схвалено науково-методичною комісією факультету комп'ютерних наук та кібернетики

Протокол від «11» лютого 2021 року № 7

Голова науково-методичної комісії  Людмила ОМЕЛЬЧУК

«11» лютого 2021 року

1. Мета дисципліни – надання необхідних знань з лінгвістики та комп'ютерної лінгвістики, їх подальше застосування для вирішенні різних задач з інтелектуальної обробки природномовних текстів.

2. Попередні вимоги до опанування або вибору навчальної дисципліни:

1. *Знати* дисципліни «Основи комп'ютерних алгоритмів», «Бази даних та інформаційні системи», «Дискретна математика».
2. *Вміти* виконувати інтелектуальний багатовимірний аналіз даних та їхню оперативну аналітичну обробку з візуалізацією результатів аналізу в процесі розв'язування прикладних задач у галузі комп'ютерних наук
3. *Володіти* мінімальними знаннями технічної англійської мови.

3. Анотація навчальної дисципліни:

Навчальна дисципліна «Основи комп'ютерної лінгвістики» є складовою вибіркового блоку «Інтелектуальні інформаційні технології» освітньо-професійної програми підготовки фахівців за першим (бакалаврським) рівнем вищої освіти *галузі знань* 12 «Інформаційні технології» *спеціальності* 122 «Комп'ютерні науки», *освітньо-професійної* програми «Комп'ютерні науки».

Дана дисципліна є вибірковою навчальною дисципліною за програмою “ Комп'ютерні науки ”. Викладається у 1 семестрі 4 курсу в обсязі – 4 кредити ECTS.

У курсі передбачено 3 частини, 3 контрольні роботи. Завершується дисципліна – іспитом в 7 семестрі.

4. Завдання (навчальні цілі):

набуття знань, умінь та навичок (компетентностей) на рівні новітніх досягнень у програмуванні, відповідно до кваліфікації фахівця з інформаційних технологій. Зокрема, розвивати здатність до алгоритмічного мислення.

5. Результати навчання за дисципліною:

| Результат навчання (1. знати; 2. вміти; 3. комунікація) | | Форми (та/або методи і технології) викладання і навчання | Методи оцінювання та пороговий критерій оцінювання (за необхідності) | Відсоток у підсумковій оцінці з дисципліни |
|--|---|--|--|--|
| Код | Результат навчання | | | |
| PH1.1 | Знати основні поняття з галузі комп'ютерної лінгвістики, закони побудови мови, базові моделі мовних явищ. | Лекція, лабораторні заняття, самостійна робота | Тест, 55% правильних відповідей | 15% |
| PH 1.2 | Знати методи аналізу природномовних тестів та відповідні алгоритми | Лекція, лабораторні заняття, самостійна робота | Тест, 55% правильних відповідей | 15% |
| PH 2.1 | Вміти аналізувати мовні явища для заданої природної мови. | Лекція, лабораторні заняття, самостійна робота | Тест, 55% правильних відповідей | 30% |
| PH 2.2 | Вміти формалізувати нестрогі методи аналізу до чітких алгоритмів та підбирати оптимальний інструментарій для виконання задачі, спираючись на машинне навчання | Лекція, лабораторні заняття, самостійна робота | Тест, 55% правильних відповідей | 30% |
| PH 3.1 | Обґрунтовувати власний погляд на задачу, спілкуватися з колегами з питань аналізу задач та проектування алгоритму | Лекція, лабораторні заняття, самостійна робота | Тест | 10% |

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

| Результати навчання дисципліни Програмні результати навчання | Результати навчання дисципліни | | | | |
|---|--------------------------------|-----|-----|-----|-----|
| | 1.1 | 1.2 | 2.1 | 2.2 | 3.1 |
| ПРН20.1. Знати і застосовувати методи інтелектуального аналізу даних та штучного інтелекту, що включають методи комп'ютерної лінгвістики та комп'ютерного зору. | | + | + | + | + |
| ПРН21.1. Знати методи машинного навчання для розв'язання прикладних задач, основні поняття та принципи роботи штучних нейронних мереж. | | | | + | + |

7. Схема формування оцінки.

7.1 Форми оцінювання студентів:

- семестрове оцінювання:

1. Контрольна робота (тест): РН 1.1.— 20 балів/11 балів.

2. Контрольна робота (тест): РН 1.2.— 20 балів/11 балів.

3. Контрольна робота (тест): РН 1.2.— 20 балів/11 балів.

- підсумкове оцінювання (у формі іспиту):

- максимальна кількість балів які можуть бути отримані студентом: 40 балів;

- результати навчання які будуть оцінюватись: РН1.1, РН1.2, РН2.1, РН2.2, РН3.1.

- форма проведення і види завдань: письмова робота.

Види завдань: 4 теоретичних та 2 практичних завдання.

Загальні вимоги:

-Студент не допускається до іспиту, якщо під час семестру набрав менше ніж 33 бали або менше 11 балів за будь-яку контрольну.

Критерії оцінювання на іспиті

| Завдання | Тема завдання | Максимальний відсоток від 40 балів | Всього відсотків |
|---------------------|--|------------------------------------|------------------|
| Завдання 1, 2, 3, 4 | Письмове запитання з наведених нижче | по 15% | 75% |
| Завдання 5 | Для заданого словника (української мови) виконати його розбір евристичним аналізом на морфемні характеристики | | |
| Завдання 6 | Для заданого набору речень англійською мовою виконати його розбір за відповідними граматиками складових та залежностей | по 25% | 25% |
| | | | 100% |

Запитання для підготовки до іспиту

1. Закон Ціпфа.
2. Синтаксичні класи.
3. Граматичні класи.
4. Контрольована мова.
5. Регулярні граматики.
6. Контекстно-вільні граматики.
7. Трансформаційні граматики.
8. Імовірнісні контекстно-вільні граматики.
9. Граматика під'єднання дерев (Tree adjoining grammar).
10. Парсинг на основі залежностей.
11. Контекстно-вільний парсинг.
12. Алгоритм Кока-Янгера-Касамі.
13. Онтології.
14. Word Net.

15. Дискурс.
16. Покрокова інтерпретація.
17. Тема і рема. Тема і фокус.
18. Теорія риторичної структури (RST).
19. Теорія центрування.
20. Анафора, її типи.
21. Розв'язання анафори на основі граматичних даних.
22. Розв'язання анафори на основі семантичних даних.
23. Алгоритм послівного напівавтоматичного поповнення словника.
24. Алгоритм групового напівавтоматичного поповнення словника.
25. Алгоритми сегментації речень.
26. Еліптичні вхідні дані.
27. Оптимальна структура для результатів парсингу.
28. Структура Link Grammar Parser.
29. Визначення значень багатозначних слів на основі WordNet.
30. Визначення значень багатозначних слів на основі СУС.
31. Метод семантичної хмари.
32. Метод синтаксичних шаблонів.
33. Нейронні мережі для семантичного аналізу.
34. Тема тексту. Способи представлення.
35. Визначення емоційної направленості тексту (sentiment analysis)
36. Автоматичне реферування.
37. Автоматичний переклад.

7.2 Організація оцінювання:

Обов'язковим є виконання модульних контрольних робіт за графіком робочої програми

Терміни проведення форм оцінювання:

1. Контрольна робота (тест): до 5 лекційного заняття.
2. Контрольна робота (тест): до 11 лекційного заняття.
3. Контрольна робота (тест): до останнього лекційного заняття.

Студент має право на одне перескладання кожної контрольної роботи із можливістю отримання максимально 80% початково визначених за цю контрольну роботу балів. Термін перескладання визначається викладачем.

7.3 Шкала відповідності оцінок

| | |
|----------------------------------|--------|
| Відмінно / Excellent | 90-100 |
| Добре / Good | 75-89 |
| Задовільно / Satisfactory | 60-74 |
| Незадовільно / Fail | 0-59 |

8. Структура навчальної дисципліни. Тематичний план лекційних занять

| № п.п. | Назва лекції | Кількість годин | | |
|---|---|-----------------|-------------|-----------------|
| | | Лекції | Лабораторії | Само-ст. роботи |
| Частина 1. Явища природної мови. Морфологія та синтаксис | | | | |
| 1. | Тема 1. Задачі комп'ютерної лінгвістики. Природні, контрольовані та штучні мови. Закон Ціпфа. Структура слова. Слова і граматичні класи. | 4 | 2 | 4 |
| 2. | Тема 2. Машинні словники, словники основані на корпусі текстів. | 2 | 2 | 4 |
| 3. | Тема 3. Звичайний синтаксис. Слова і синтаксичні класи. Природні граматики та мови. Формальні граматики та мови. | 2 | | 4 |
| 4. | Тема 4. Регулярні граматики. Контекстно-вільні граматики. Трансформаційні граматики. | 2 | 2 | 4 |
| 5. | <i>Контрольна робота 1</i> | 2 | | |
| Частина 2. Значення слів, онтології. Дискурс. Значення тексту. | | | | |
| 6. | Тема 5. Онтології. Фреймові ієрархічні моделі. | 4 | | 4 |
| 7. | Тема 6. WordNet. СУС. | 2 | | 4 |
| 8. | Тема 7. Автоматичне визначення значень багатозначних слів. | 2 | 2 | 4 |
| 9. | Тема 8. Автоматичне членування тексту. Виділення термінів. | 2 | 2 | 4 |
| 10. | Тема 9. Дискурс. Покрокова інтерпретація. Тема і рема. Тема і фокус. | 2 | | 4 |
| 11. | Тема 10. Теорія риторичної структури (RST). Теорія центрування. | 2 | 2 | 4 |
| 12. | <i>Контрольна робота 2</i> | 2 | | |
| Частина 3. Етапи обробки тексту. | | | | |
| 13. | Тема 11. Визначення теми тексту. Автоматична кластеризація потоків текстових повідомлень. Визначення спаму. | 4 | | 4 |
| 14. | Тема 12. Визначення емоційної направленості тексту (sentiment analysis) | 4 | 2 | 6 |
| 15. | Тема 13. Автоматичне реферування. | 2 | | 6 |
| 16. | Тема 14. Автоматичний переклад на основі імовірнісних перетворень. | 2 | | 6 |
| 17. | <i>Контрольна робота 3</i> | 2 | | |
| | ВСЬОГО | 42 | 14 | 62 |

Загальний обсяг 120 год., в тому числі:

Лекції – 42 год.

Консультації - 2 год.

Самостійна робота - 62 год.

9. Рекомендовані джерела:

Основні:

1. Волошин В.Г. Комп'ютерна лінгвістика: Навчальний посібник. – Суми: Університетська книга, 2004. –382 с.
 2. Анисимов А.В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык Киев: Наук. думка, 1988.- 223 с.
 3. Партико З.В. Прикладна і комп'ютерна лінгвістика, Львів, «Афіша», 2008, - 221 с.
- В тому числі й інтернет ресурси*
4. The Oxford handbook of computational linguistics R. Mitkov (Ed) Oxford University Press Ел. ресурс. Режим доступу: http://books.google.com.ua/books/about/The_Oxford_handbook_of_computational_lin.html?id=OaClhre-vW4C&redir_esc=y
 5. Daniel Jurafsky and James H. Martin Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics. 2nd edition, 2009 Ел. ресурс. Режим доступу: http://rapidlibrary.com/files/speech-and-language-processing-an-introduction-to-natural-language-processing-djv_23726564.html

Додаткові:

6. Український правопис / Ін-т мовознавства ім. О.О. Потебні НАН України, Ін-т укр. мови НАН України. — К. : Наук. думка, 2007. — 288 с.
 7. The Oxford Handbook of Applied Linguistics / Ed. by R.Kaplan. NY: Oxford university press, 2002, 672 P.
 8. J. Allen, “Natural language understanding” Menlo Park, Calif. Benjamin/Cummings 1995, 654 P.
- В тому числі й інтернет ресурси*
9. Сайт проекту WordNet Ел. ресурс. Режим доступу <http://wordnet.princeton.edu/>
 10. Спільний сайт Інституту російсько мови ім. В.В. Виноградова та компанії «СЛОВАРИ.РУ» Ел. ресурс. Режим доступу <http://www.slovari.ru/>