

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
Кафедра оптоелектроніки та інформаційних технологій

Затверджено

На засіданні кафедри оптоелектроніки та
інформаційних технологій
факультету електроніки та комп'ютерних
технологій
Львівського національного університету імені
Івана Франка
(протокол №8 від 31 серпня 2020 р.)

Завідувач кафедри



Силабус з навчальної дисципліни
«Комп'ютерна лінгвістика та лексикографічні системи»,
що викладається в межах ОПП «Комп'ютерні науки»
третього рівня вищої освіти (доктор філософії)
для здобувачів з спеціальності
122 «Комп'ютерні науки»

| | |
|--|--|
| Назва дисципліни | Комп'ютерна лінгвістика та лексикографічні системи |
| Адреса викладання дисципліни | м. Львів, вул. Тарнавського, 107 |
| Факультет та кафедра, за якою закріплена дисципліна | Факультет електроніки та комп'ютерних технологій, кафедра оптоелектроніки та інформаційних технологій |
| Галузь знань, шифр та назва спеціальності | 12 Інформаційні технології, 122 Комп'ютерні науки |
| Викладачі дисципліни | Кушнір Олег Степанович, докт. фіз.-мат. наук, проф., проф. |
| Контактна інформація викладачів | oleh.kushnir@lnu.edu.ua https://electronics.lnu.edu.ua/employee/kushnir-o-s |
| Консультації з питань навчання по дисципліні відбуваються | Консультації в день проведення лекційних занять (за попередньою домовленістю). Також можливі онлайн-консультації через Zoom. Для погодження часу онлайн-консультацій слід писати на електронну пошту викладача. |
| Сторінка дисципліни | https://electronics.lnu.edu.ua/course/komp-iuterna-linhvistyka-its http://194.44.208.156/moodle/course/view.php?id=59 |
| Інформація про дисципліну | Дисципліна «Комп'ютерна лінгвістика та лексикографічні системи» є дисципліною вільного вибору аспірантів зі спеціальності 122 Комп'ютерні науки для освітньої програми «Комп'ютерні науки», яка викладається в 4 семестрі в обсязі 3,0 кредитів (за Європейською Кредитно-Трансферною Системою – ECTS). |
| Коротка анотація дисципліни | Навчальну дисципліну розроблено для одержання аспірантами теоретичних знань з комп'ютерної лінгвістики та лексикографічних систем, а також для формування в них навичок ефективного застосування засвоєних знань і методів у розв'язанні прикладних задач комп'ютерної лінгвістики. Представлено теоретичні основи комп'ютерної, статистичної та математичної лінгвістики, класифікація та огляд особливостей відомих продуктів комп'ютерної лінгвістики, основи машинного перекладу та комп'ютерної лексикографії, а також відповідні комп'ютерні алгоритми і засоби обробки даних. Описано складні системи та мережі, пов'язані з предметом розгляду. |
| Мета та цілі дисципліни | Метою вивчення дисципліни «Комп'ютерна лінгвістика та лексикографічні системи» є ознайомлення аспірантів з теоретичними основами комп'ютерної лінгвістики, а також формування в аспірантів практичних навичок, які б дали змогу ефективно застосовувати засвоєні знання, алгоритми, методи та прикладні програми комп'ютерної лінгвістики. |
| Література для вивчення дисципліни | Основна література: 1. Bolshakov I. Computational linguistics. Models, resources, applications / I. Bolshakov, A. Gelbukh. – Mexico : Ciencia de la Computacion, 2004. – 198 p. 2. Волошин В. Г. Комп'ютерна лінгвістика / В. Г. Волошин. – Суми : Університетська книга, 2004. – 382 с. 3. Шемакин Ю. И. Начала компьютерной лингвистики / Ю. И. Шемакин. – Москва : Изд-во МГОУ, 1992. – 182 с. 4. Баранов А. Н. Введение в прикладную лингвистику / А. Н. Баранов. – Москва : Эдиториал УРСС. – 2001. – 360 с. 5. Мирам Г. Алгоритмы перевода: Вступительный курс по формализации перевода / Г. Мирам. – Киев : Эльга, Ника-Центр, 2004. – 176 с. 6. Хархалис Р. И. Компьютерный перевод иностранных текстов / Р. И. Хархалис. Киев : Терези, 1998. – 193 с. 7. Белоногов Г. Г. Компьютерная лингвистика и перспективные информационные технологии / Г. Г. Белоногов. – Москва : Русский |

| | |
|---|---|
| | <p>мир, 2004. – 248 с.</p> <p>8. Большакова Е. И. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / Е. И. Большакова, Э. С. Клышинский, Д. В. Ландэ, А. А. Носков, О. В. Пескова, Е. В. Ягунова. – Москва : МИЭМ, 2011. – 272 с.</p> <p>9. https://www.gutenberg.org/ Додаткова література:</p> <p>10. Пиотровский Р. Г. Математическая лингвистика / Р. Г. Пиотровский, К. Б. Бектаев, А. А. Пиотровский. – Москва : Высшая школа, 1977. – 383 с.</p> <p>11. Фланаган Дж. Л. Анализ, синтез и восприятие речи / Дж. Л. Фланаган. – Москва : Связь, 1968. – 398 с.</p> <p>12. Маркел Дж. Д. Линейное предсказание речи / Дж. Д. Маркел, А. Х. Грэй. – Москва : Связь, 1980. – 308 с.</p> <p>13. Рассел С. Искусственный интеллект. Современный подход / С. Рассел, П. Норвинг. – Москва : Изд. дом «Вильямс», 2006. – 1408 с.</p> <p>14. Анисимов А. В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык / А. В. Анисимов. – Киев : Наукова думка, 1991. – 208 с.</p> <p>15. 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.).</p> |
| Обсяг курсу | Сумарно 90 годин. Із них 32 години лекцій, 16 годин лабораторних робіт і 42 годин самостійної роботи |
| Очікувані результати навчання | <p>Після завершення цього курсу аспірант буде:</p> <ul style="list-style-type: none"> - знати основні методи комп'ютерної лінгвістики, основні теорії, моделі та алгоритми комп'ютерної лінгвістики і опису лінгвістичних систем, інформаційного пошуку та інтелектуального аналізу текстових даних; - вміти аналізувати моделі комп'ютерної лінгвістики, працювати з відповідними програмними продуктами, застосовувати комп'ютерну техніку для вирішення лінгвістичних задач, розробляти та реалізувати відповідні алгоритми, писати прикладні програми та користуватися ними. |
| Ключові слова | Комп'ютерна лінгвістика, лексикографічні системи статистична лінгвістика, машинний переклад, комп'ютерна лексикографія |
| Формат курсу | Очний |
| | Проведення лекцій, лабораторних робіт та консультації для поглибленого розуміння тем |
| Теми | Див. СХЕМА КУРСУ |
| Підсумковий контроль, форма | Іспит вкінці семестру |
| Пререквізити | Для вивчення курсу аспіранти потребують базових знань з дисциплін «Вища математика», «Дискретна математика», «Алгоритми та структури даних», «Чисельні методи», «Теорія ймовірності та математична статистика», «Об'єктно-орієнтоване програмування», «Бази даних та знань». |
| Навчальні методи та техніки, які будуть використовуватися під час викладання курсу | Лекції, презентації, лабораторні роботи, індивідуальні практичні завдання обговорення, дискусії. |
| Необхідне обладнання | Мультимедіа, платформи Moodle і Zoom, комп'ютерне програмне забезпечення |
| Критерії оцінювання | Оцінювання проводиться упродовж семестру та під час екзаменаційної |

| | |
|---|---|
| <p>(окремо для кожного виду навчальної діяльності)</p> | <p>сесії за 100-бальною шкалою. Бали нараховуються за такими видами робіт із таким співвідношенням:</p> <ul style="list-style-type: none"> • лабораторні роботи: 40% оцінки; максимальна кількість балів 40. • контрольні заміри (2 модулі): 10% оцінки; максимальна кількість балів 10. • іспит: 50% оцінки; максимальна кількість балів 50. <p>Загалом 100 балів.</p> <hr/> <p>Контрольні заміри знань проводять у формі стандартних практичних завдань і теоретичних питань.</p> <p>Академічна доброчесність: Очікується, що лабораторні та контрольні роботи аспірантів будуть їхніми оригінальними дослідженнями або міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших аспірантів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в роботі аспіранта є підставою для її незарахування викладачем, незалежно від масштабів плагіату або спроб обману.</p> <p>Відвідування занять є важливою складовою навчання. Очікується, що всі аспіранти відвідають усі лекції та лабораторні заняття курсу. Аспіранти мають інформувати викладача про неможливість відвідати заняття. Аспіранти зобов'язані дотримуватися всіх термінів, визначених для виконання видів робіт, передбачених курсом.</p> <p>Література. Уся література, яку аспіранти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Аспіранти також заохочуються до використання іншої літератури та джерел, зокрема наукової літератури, яка відсутня серед обов'язкової та рекомендованої.</p> <p>Політика виставлення балів. Враховуються бали, набрані на поточному опитуванні, самостійній роботі та бали підсумкового контролю знань. Обов'язково враховуються присутність на заняттях та активність аспіранта під час лабораторних занять; наголошується на неприпустимості пропусків або запізень на заняття, користування мобільним телефоном, планшетом або іншими мобільними пристроями під час занять з метою, не пов'язаною з навчанням, списування та плагіату, несвоєчасного виконання поставлених завдань і т. ін.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p> |
| <p>Питання до контрольних робіт</p> | <p>Перелік питань і завдань для проведення підсумкової оцінки знань усіх тем курсу до контрольних робіт розміщено на веб-сторінці.</p> |
| <p>Опитування</p> | <p>Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.</p> |

СХЕМА КУРСУ

| Тиж. | Тема, план, короткі тези | Форма діяльності (заняття) | Література. Ресурси в Інтернеті | Завдання (лабораторна робота), год. | Термін виконання |
|------|--|----------------------------|---------------------------------|---------------------------------------|------------------|
| 1 | <p>Вступ. Лінгвістика та її структура. 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.</p> | Лекція | 2, 3, 7 | Класифікація та кластеризація текстів | 2 тиж. семестру |
| 2, 3 | <p>Основні поняття лінгвістики.</p> | Лекція | 2, 4, 7 | Програмування задачі | 4 тиж. |

| | | | | | |
|--------|--|--------|---------------------|--|------------------|
| | Загальні поняття про мову. Мова і мислення. Фонетика. Морфологія. Синтаксис. Семантика. Семантичні мережі. Прагматика. | | | визначення співвідношення кількості голосних і приголосних (прозорості мови) для українських текстів | семестру |
| 4 | Розвиток ідей, теорій, підходів і методів комп'ютерної лінгвістики. Структурний підхід. Внесок Н. Хомського. Контекстно-вільна граматики. Трансформаційні граматики. Валентності. Теорія ЗМІСТ ↔ ТЕКСТ. Моделі керування. Дерева залежностей. Статистичні підходи. | Лекція | 1, 2, 8 | Програмування задач морфологічного синтезу дієслівних форм | 5 тиж. семестру |
| 5, 6 | Продукти комп'ютерної лінгвістики: досягнення та проблеми. Класифікація прикладних лінгвістичних систем. Розстановка переносів. Перевірка орфографії, граматики та стилістики. Інформаційні пошукові системи. Інтерфейс на природній мові. Видобування фактичних даних. Генерація тексту. | Лекція | 1, 4, 8, 9 | Робота з базами лінгвістичних даних. Побудова запитів з урахуванням лінгвістичних міркувань | 7 тиж. семестру |
| 7 | Мова як двонаправлений перетворювач змісту та тексту. Поняття тексту і змісту. Способи представлення змісту. Розкладання і атомізація змісту. Неоднозначність “картування”. Багатоетапність перетворення ЗМІСТ ↔ ТЕКСТ. | Лекція | 2, 4, 7 | Визначення ключових слів і змістового навантаження текстів | 8 тиж. семестру |
| 8, 9 | Лінгвістичні знаки. Лінгвістичний знак у теорії ЗМІСТ ↔ ТЕКСТ і граматики структури фраз. Порівняння генеративних ідей та ідеї обмежень. | Лекція | 1, 2, 10 | Програмування опрацювання лінгвістичних одиниць | 10 тиж. семестру |
| 10, 11 | Лінгвістичні моделі. Нейролінгвістичні, психолінгвістичні та функціональні моделі мови. Спільні риси сучасних моделей. Модель ЗМІСТ ↔ ТЕКСТ. Зредуковані моделі. Аналогії в природних мовах. Емпіричні та раціоналістичні підходи. Статистичні моделі. | Лекція | 1, 3, 4, 7, 10 | Розрахунок статистичних параметрів текстів | 12 тиж. семестру |
| 12, 13 | Автоматичне введення звуків мови, аналіз та розпізнавання мови комп'ютером. Акустична будова звуків. Принципи мовної комунікації. Аналого-цифрове перетворення мовних сигналів. Автоматичне розпізнавання. Звуковий спектрограф. Спектрально-часовий аналіз. Бінарна селекція звукових елементів. Методика | Лекція | 1, 7, 8, 11, 12, 13 | Вирішення задач розпізнавання за допомогою пакету FineReader | 14 тиж. семестру |

| | | | | | |
|--------|--|--------|----------------|--|------------------|
| | формантного аналізу мови. | | | | |
| 14, 15 | Синтез мови та мовні технології. Розвиток систем синтезу мовлення. Кодування сигналів. Формантні синтезатори. Предикативне кодування. Алгоритм синтезу. Структура та види синтезаторів. | Лекція | 7, 8, 11, 12 | Обробка природних мов у рамках середовища Python | 15 тиж. семестру |
| 16 | Машинний переклад і комп'ютерна лексикографія. Оцінка та критерії якості. Типові помилки. Теоретичні моделі. Алгоритми аналізу та синтезу. Види перекладу. Сучасний стан. Проблеми комп'ютерної лексикографії. | Лекція | 1, 5, 6, 8, 13 | Програмний пакет Stylus: налаштування інтерфейсу та робота | 16 тиж. семестру |