



ОСНОВИ КОРПУСНОЇ ТА КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ

Силабус –

Робоча програма навчальної дисципліни (освітнього компонента)

Реквізити навчальної дисципліни

| | |
|---|--|
| Рівень вищої освіти | <i>Другий (магістерський)</i> |
| Галузь знань | <i>03 Гуманітарні науки</i> |
| Спеціальність | <i>035 Філологія</i> |
| Освітня програма | <i>Германські мови та літератури (переклад включно), перша — німецька</i> |
| Статус дисципліни | <i>Вибіркова</i> |
| Форма навчання | <i>Очна (денна)</i> |
| Рік підготовки, семестр | <i>1 курс, весняний семестр</i> |
| Обсяг дисципліни | <i>120 год (4 кредити): практичні заняття - 36 год., індивідуальні заняття – 18 год., СРС – 66 год.</i> |
| Семестровий контроль/ контрольні заходи | <i>Залік/модульна контрольна робота</i> |
| Розклад занять | <i>http://rozklad.kpi.ua/</i> |
| Мова викладання | <i>Німецька та українська</i> |
| Інформація про керівника курсу / викладачів | <i>Практичні: к.філол.н, доцент Івашкевич Леся Сергіївна, lesia.ivashkevych@gmail.com</i> |
| Розміщення курсу | <i>https://do.ipk.kpi.ua/user/files.php</i> |

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Предмет вивчення — корпусні та комп'ютерні методи обробки природної мови.

Метою навчальної дисципліни є формування у студентів здатностей:

- будувати алгоритми для виявлення, формулювання та вирішення проблем у лінгвістичних дослідженнях на належному науковому та методологічному рівні;
- проводити пошук, опрацювання та аналіз інформації з різних професійних джерел;
- працювати в команді та автономно;
- мислити абстрактно, здійснювати аналіз, синтез та інтерпретацію отриманих даних;
- орієнтуватися у перевагах та обмеженнях корпусних досліджень та змінах у дослідницьких парадигмах, які зумовлює корпусна лінгвістика;
- вільно оперувати термінологією корпусної та комп'ютерної лінгвістики;
- реалізувати системний підхід до організації та ефективного здійснення емпіричного лінгвістичного дослідження;
- розробляти структурно-сміслові алгоритми для проведення таких досліджень та реалізації поставлених завдань.

У результати вивчення дисципліни “Основи корпусної та комп’ютерної лінгвістики” студенти зможуть:

- застосовувати сучасні методики і технології, зокрема інформаційні, для успішного й ефективного здійснення професійної науково-інноваційної діяльності;
- порівнювати корпусні методи досліджень із традиційними;
- збирати й систематизувати мовні та мовленнєві факти;
- здійснювати науковий аналіз мовного та мовленнєвого матеріалу, інтерпретувати та структурувати його з урахуванням корпусних методологічних принципів, формулювати узагальнення на основі самостійно проведеного аналізу даних;
- вибирати оптимальні дослідницькі підходи й методи для аналізу конкретного лінгвістичного матеріалу;
- планувати, організовувати, здійснювати і презентувати лінгвістичні дослідження.

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

Дисципліна “Основи корпусної та комп’ютерної лінгвістики” вивчається протягом 2 семестру. Необхідною передумовою є вивчення дисциплін «Практикум з мовної комунікації та перекладу (німецька мова)» ПО 1 та “Наукова робота за темою магістерської дисертації” ПО 6. Отримані під час вивчення дисципліни знання та вміння забезпечують успішне написання та захист магістерської дисертації. Для успішного опанування дисципліною необхідно вміти працювати з комп’ютером як впевнений користувач.

3. Зміст навчальної дисципліни

Розділ 1. Основні поняття корпусної та комп’ютерної лінгвістики.

Тема 1. Історія становлення та основи сучасної корпусної лінгвістики. Вступ в корпусну лінгвістику. Основні терміни і поняття. Корпус, анотація, сегментування, маркування.

Тема 2. Типологія корпусів. Огляд наявних корпусів різних типів.

Тема 3. Можливості використання різних типів корпусів у лінгвістичних дослідженнях. Характерні риси корпусного дослідження.

Тема 4. Можливості використання різних типів корпусів у лінгвістичних дослідженнях. Характерні риси корпусного дослідження.

Тема 5. Створення власного корпусу: вимоги, формати, метадані.

Тема 6. Ефективний пошук по корпусу за допомогою регулярних виразів. Основи.

Тема 7. Ефективний пошук по корпусу за допомогою регулярних виразів. Закріплення.

Тема 8. Практичні завдання на пошук за допомогою регулярних виразів.

Тема 9. Робота з корпусним менеджером. Основні функції — конкорданс, перелік частотності слів.

Тема 10. Робота з корпусним менеджером. Основні функції — ключові слова тексту, колокації, N-грами.

Тема 11. Робота з корпусним менеджером. Пошук з активованою функцією регулярних виразів. Експорт результатів.

Тема 12. Можливості застосування корпусних методів дослідження у написанні магістерської дисертації.

Розділ 2. Програмування для обробки природної мови.

Тема 1. Основи програмування на Python. Стрічки, змінні, функція print.

Тема 2. Основи програмування на Python. Помилки синтаксису. Списки.

Тема 3. Основи програмування на Python. Операції зі списками. Методи split, strip.

Тема 4. Основи програмування на Python. Цикли. Їх застосування для обробки текстів. Інкрементація.

Тема 5. Основи програмування на Python. Підключення модуля регулярних виразів. Вирішення практичних задач.

4. Навчальні матеріали та ресурси

Базова література:

1. Bubenhofer, Noah (2006-2017): Einführung in die Korpuslinguistik: Praktische Grundlagen und Werkzeuge. Elektronische Ressource: <http://www.bubenhofer.com/korpuslinguistik/>.

Додаткова література:

2. Рыков В.В. Курс лекций по корпусной лингвистике. [Электронный ресурс]. – Режим доступа: <http://rykov-cl.narod.ru/c.html>.

3. Reinhard Köhler. Korpuslinguistik – zu wissenschaftstheoretischen Grundlagen und methodologischen Perspektiven. LDV-Forum 2005 – Band 20 (2) – S. 1-16.

4. Жуковська В.В. Вступ до корпусної лінгвістики. Навчальний посібник. — Житомир: Вид-во ЖДУ ім. І. Франка, 2013. — 142 с.

5. Scherer, Carmen. Korpuslinguistik. — Heidelberg: Universitätsverlag Winter, 2014.

6. Lemnitzer, Lothar und Zinsmeister, Heike. Korpuslinguistik. Eine Einführung (3. Auflage) — Tübingen: Narr., 2010.

7. Perkuhn, Rainer; Keibel, Holger; Kupietz, Marc. Korpuslinguistik. —Paderborn: Fink, 2012.

8. <https://regexr.com>

9. <https://regexone.com/>

10. <https://repl.it>

Навчальний контент

5. Методика опанування навчальної дисципліни:

У курсі інтенсивно використовуються практичні методи навчання, базовані на самостійних діях студентів: виконанні завдань, підготовці власних кейсів, роботі з тьюторіалами та інтерактивними середовищами для тренування практичних навиків. Широко використовуються засоби унаочнення: презентації, аудіо- та відеоформати подачі інформації.

Для активізації самостійного мислення студентів та вироблення власної позиції з актуальних питань дисципліни рекомендовано використовувати групові форми роботи з наступним загальним обговоренням під модерацією викладача.

Робота в межах індивідуальних занять здійснюється у форматі консультацій.

Календарний план

| № п/п | Зміст навчальної роботи | Практичні заняття (год.) | Робота на індивідуальних заняттях |
|-------|---|--------------------------|-----------------------------------|
| | Розділ 1. | | |
| 1 | Історія становлення та основи сучасної корпусної лінгвістики. Вступ в корпусну лінгвістику. Основні терміни і поняття. Корпус, анотація, сегментування, маркування. | 2 | |
| 2 | Типологія корпусів. Огляд наявних корпусів різних типів. | 2 | |

| № п/п | Зміст навчальної роботи | Практичні заняття (год.) | Робота на індивідуальних заняттях |
|-------|--|--------------------------|--|
| 3 | Можливості використання різних типів корпусів у лінгвістичних дослідженнях. Характерні риси корпусного дослідження. | 2 | |
| 4 | Вимоги до корпусного дослідження. Веб як корпус: переваги та недоліки. | 2 | Консультування щодо створення власного корпусу (6 год.) |
| 5 | Створення власного корпусу: вимоги, формати, метадані. | 2 | |
| 6 | Ефективний пошук по корпусу за допомогою регулярних виразів. Основи. | 2 | |
| 7 | Ефективний пошук по корпусу за допомогою регулярних виразів. Закріплення. | 2 | |
| 8 | Практичні завдання на пошук за допомогою регулярних виразів. Проміжне тестування 1. | 2 | |
| 9 | Робота з корпусним менеджером. Основні функції — конкорданс, перелік частотності слів. | 2 | Консультування з використання корпусів для магістерських дисертацій (8 год.) |
| 10 | Робота з корпусним менеджером. Основні функції — ключові слова тексту, колокації, N-грами. | 2 | |
| 11 | Робота з корпусним менеджером. Пошук з активованою функцією регулярних виразів. Експорт результатів. | 2 | |
| 12 | Можливості застосування корпусних методів дослідження у написанні магістерської дисертації. | 2 | |
| | Розділ 2. | | |
| 13 | Основи програмування на Python. Стрічки, змінні, функція print. | 2 | Консультування про проходження курсу програмування з Python на освітніх платформах. (4 год.) |
| 14 | Основи програмування на Python. Помилки синтаксису. Списки. Проміжне тестування 2. | 2 | |
| 15 | Основи програмування на Python. Операції зі списками. Методи split, strip. | 2 | |
| 16 | Основи програмування на Python. Цикли. Їх застосування для обробки текстів. Інкрементація. | 2 | |
| 17 | Основи програмування на Python. Підключення модуля регулярних виразів. Вирішення практичних задач. Модульна контрольна робота. | 2 | |
| 18 | Повторення. Підвищення рейтингу. Залік. | 2 | |

6. Самостійна робота студента

| № п/п | Зміст навчальної роботи | Рекомендований час СРС |
|-------|---|------------------------|
| | Розділ 1. | |
| 1 | Історія становлення та основи сучасної корпусної лінгвістики. Вступ в корпусну лінгвістику. Основні терміни і поняття. Корпус, анотація, сегментування, маркування. Самостійна робота: ознайомитися з основами корпусної лінгвістики на сайті Ноа Бубнгофера. | 2 |
| 2 | Типологія корпусів. Огляд наявних корпусів різних типів. Самостійна робота: практична робота з корпусами німецької, англійської та української мов. | 4 |
| 3 | Можливості використання різних типів корпусів у лінгвістичних дослідженнях. Характерні риси корпусного дослідження. Самостійна робота: власний пошук корпуснобазованих досліджень серед мовознавчих досліджень в Україні та закордоном. | 4 |
| 4 | Вимоги до корпусного дослідження. Веб як корпус: переваги та недоліки. Самостійна робота: опрацювати статтю К. Штаті про ознаки корпусного дослідження. | 2 |
| 5 | Створення власного корпусу: вимоги, формати, метадані. Самостійна робота: ознайомитися з алгоритмом створення власного корпусу на сайті Ноа Бубенгофера. | 4 |
| 6 | Ефективний пошук по корпусу за допомогою регулярних виразів. Основи. Самостійна робота: опрацювати перші 10 кроків за онлайн-тьюторіалом regexone.com | 2 |
| 7 | Ефективний пошук по корпусу за допомогою регулярних виразів. Закріплення. Самостійна робота: вирішення практичних завдань з регулярних виразів у редакторі regex.com . розробка власних завдань один для одного на роботу з регулярними виразами. | 4 |
| 8 | Практичні завдання на пошук за допомогою регулярних виразів. Самостійна робота: індивідуальна робота з вирішення практичних завдань з пошуку регулярними виразами. | 4 |
| 9 | Робота з корпусним менеджером. Основні функції — конкорданс, перелік частотності слів. | 2 |
| 10 | Робота з корпусним менеджером. Основні функції — ключові слова тексту, колокації, N-грами. Самостійна робота: практична робота з пошуку у корпусному менеджері по власному корпусі. | 2 |

| № п/п | Зміст навчальної роботи | Рекомендований час СРС |
|-------|--|------------------------|
| 11 | Робота з корпусним менеджером. Пошук з активованою функцією регулярних виразів. Експорт результатів. Самостійна робота: практична робота з пошуку у корпусному менеджері по власному корпусі з використанням пошуку регулярними виразами. | 4 |
| 12 | Можливості застосування корпусних методів дослідження у написанні магістерської дисертації. Самостійна робота: підготовка презентації на тему можливостей застосування корпусного дослідження під час роботи над власною магістерською дисертацією. | 4 |
| | Розділ 2. | |
| 13 | Основи програмування на Python. Стрічки, змінні, функція print. Самостійна робота: вирішення практичних завдань в інтерактивному середовищі repl.it. | 4 |
| 14 | Основи програмування на Python. Помилки синтаксису. Списки. Самостійна робота: вирішення практичних завдань в інтерактивному середовищі repl.it. | 4 |
| 15 | Основи програмування на Python. Операції зі списками. Методи split, strip. Самостійна робота: вирішення практичних завдань в інтерактивному середовищі repl.it. | 4 |
| 16 | Основи програмування на Python. Цикли. Їх застосування для обробки текстів. Інкрементація. Самостійна робота: вирішення практичних завдань в інтерактивному середовищі repl.it. | 4 |
| 17 | Основи програмування на Python. Підключення модуля регулярних виразів. Вирішення практичних задач. Самостійна робота: вирішення практичних завдань в інтерактивному середовищі repl.it. Підготовка до МКР. | 6 |
| 18 | Повторення. Підвищення рейтингу. | 6 |

Політика та контроль

7. Політика навчальної дисципліни

На заняттях студенти працюють з інтернет-ресурсами, в тому числі англomовними, тому для успішного засвоєння курсу потрібні знання англійської на рівні B1 та готовність працювати з матеріалами інструктивного характеру (тьюторіалами та відеокурсами, інтерактивними середовищами) та новим для них програмним забезпеченням.

Для успішного проходження курсу важливо працювати синхронно, не відстаючи від групи.

Завдання в межах індивідуальних занять надсилаються викладачу електронною поштою або через мережу Telegram згідно встановлених умовами проєкту дедлайнів. На консультаціях (дати

узгоджуються заздалегідь) студенти ставлять запитання та отримують допомогу щодо самостійної роботи.

Система оцінювання не передбачає заохочувальних та штрафних балів.

Пропущені контрольні заходи

Кожен студент має право відпрацювати пропущені з поважної причини (лікарняний, мобільність тощо) заняття за рахунок самостійної роботи.

Детальніше за посиланням: https://document.kpi.ua/files/2020_7-137.pdf.

Академічна доброчесність

Студенти зобов'язані дотримуватися правил академічної доброчесності. Політику та принципи академічної доброчесності визначено в розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: https://me.kpi.ua/downloads/Kodeks_chesti.pdf.

Норми етичної поведінки

Норми етичної поведінки студентів і працівників визначено в розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: https://me.kpi.ua/downloads/Kodeks_chesti.pdf.

Процедура оскарження результатів контрольних заходів

Студенти мають можливість підняти будь-яке питання, яке стосується процедури контрольних заходів та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами. Студенти мають право оскаржити результати контрольних заходів, але обов'язково аргументовано, пояснивши з яким критерієм не погоджуються. Детальніше: https://document.kpi.ua/files/2020_7-137.pdf.

Інклюзивне навчання

Дисципліна «Основи корпусної та комп'ютерної лінгвістики» може викладатися для більшості студентів з особливими освітніми потребами.

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

Рейтинг студента складається з балів, які він отримує за такі контрольні заходи:

| № з/п | Контрольний захід | % | Ваговий бал |
|-------|-----------------------|----|-------------|
| 1. | Проміжне тестування 1 | 30 | 30 |
| 2. | Проміжне тестування 2 | 30 | 30 |
| 3. | МКР | 40 | 40 |
| | Всього | | 100 |

Основні критерії оцінювання роботи студентів

1. Проміжне тестування 1 — 30 балів

Проміжне тестування складається із 30 запитань з варіантами відповідей, кожне з яких оцінюється максимум в 1 бал.

Критерії оцінювання:

1 бал — відповідь правильна;

0 балів — відповідь неправильна.

- «відмінно» – правильні відповіді на не менше ніж на 90% запитань – 30-27 балів;
- «добре» – правильні відповіді на не менше ніж на 75% запитань – 26-22 бали;

- «задовільно» – правильні відповіді на не менше ніж на 60% запитань – 21-18 балів;
- «незадовільно» – відповідь не відповідає вимогам до «задовільно» – 0 балів.

2. Проміжне тестування 2 — 30 балів

Проміжне тестування складається із 30 запитань з варіантами відповідей, кожне з яких оцінюється максимум в 1 бал.

Критерії оцінювання:

1 бал — відповідь правильна;
0 балів — відповідь неправильна.

- «відмінно» – правильні відповіді на не менше ніж на 90% запитань – 30-27 балів;
- «добре» – правильні відповіді на не менше ніж на 75% запитань – 26-22 бали;
- «задовільно» – правильні відповіді на не менше ніж на 60% запитань – 21-18 балів;
- «незадовільно» – відповідь не відповідає вимогам до «задовільно» – 0 балів.

3. МКР — 40 балів

Модульна контрольна робота проводиться у форматі комплексного фінального тестування. Зміст МКР складають 40 запитань з варіантами відповідей, правильна відповідь на кожне запитання оцінюється в 1 бал.

Критерії оцінювання:

1 бал — відповідь правильна;
0 балів — відповідь неправильна.

- «відмінно» – правильні відповіді на не менше ніж на 90% запитань – 40-36 балів;
- «добре» – правильні відповіді на не менше ніж на 75% запитань – 35-30 балів;
- «задовільно» – правильні відповіді на не менше ніж на 60% запитань – 29-24 бали;
- «незадовільно» – відповідь не відповідає вимогам до «задовільно» – 0 балів.

Календарний контроль:

| Критерій | | Перший календарний контроль | Другий календарний контроль |
|------------------------------|------------------|-----------------------------|-----------------------------|
| Термін календарного контролю | | 8-ий тиждень | 14-ий тиждень |
| Умови: | Поточний рейтинг | ≥ 15 балів | ≥ 30 балів |

Семестровий контроль:

| Отримання заліку | | Критерій |
|------------------|------------------|----------|
| 1 | Поточний рейтинг | RD ≥ 60 |

Здобувачі, які виконали всі умови допуску до заліку та мають рейтингову оцінку 60 і більше балів, отримують відповідну до набраного рейтингу оцінку без додаткових випробувань. Студенти, які наприкінці семестру мають рейтинг менш 60 балів, а також ті, хто хоче підвищити оцінку, виконують залікову роботу і ця рейтингова оцінка є остаточною. Контрольна залікова робота полягає у виконанні 5 практичних завдань за матеріалами курсу. У разі написання залікової контрольної роботи на оцінку, нижчу за отриману автоматом, попередній рейтинг з дисципліни скасовується і до залікової відомості заноситься сума балів за залікову контрольну роботу («жорстка» PCO).

Залікова контрольна робота оцінюється у 100 балів максимум.

Критерії оцінювання кожного завдання:

«Відмінно»: 20-18 балів — завдання виконано правильно та логічно, використані ефективні засоби та інструменти. Студент може за потреби пояснити хід виконання завдання.

«Добре»: 17-15 балів — загалом завдання виконано правильно, використані ефективні засоби та інструменти, однак допущено кілька неточностей. Студент може за потреби пояснити хід виконання завдання.

«Задовільно»: 14-12 балів — загалом завдання виконано правильно, використані ефективні засоби та інструменти, однак допущено кілька неточностей. У студента виникають труднощі під час пояснення виконання завдань.

«Незадовільно»: 0 балів — завдання виконано неправильно або ж студент не може пояснити хід його виконання.

Умовою допуску до семестрового контролю є:

- зараховані проміжні тестування (мінімум 36 б.);

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

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

9. Додаткова інформація з дисципліни:

9.1. На контроль виносяться такі орієнтовні питання:

1. Скласти покрокову схему методології корпусного дослідження, пояснити кожен крок.
2. Написати формули регулярних виразів, якими можна спіймати в документі:
 - номери телефонів у різних форматах;
 - електронні адреси у різних форматах;
 - слова, довші за 12 літер;
 - комбінації “прикметник+іменник”;
3. Проаналізувати запропоновані теми на предмет доцільності їх дослідження корпусними методами.
4. Показати функції та можливості українського корпусу `moval.info`.
5. Показати функції та можливості українського корпусу ГРАК.
6. Показати функції та можливості німецького корпусу DWDS.
7. Показати функції та можливості німецького корпусу COSMAS.
8. Показати основні функції корпусного менеджера AntConc.
9. Прокоментувати роботу парсера ParZu.
10. Вказати помилки у програмах на Python.
11. Написати прості програми на Python за пройденими темами.

9.2. В межах дисципліни можуть бути зараховані результати навчання, отримані під час проходження курсів з корпусної та комп'ютерної лінгвістики на відкритих освітніх платформах за умови наявності відповідного підтвердження. Детальніше: <https://osvita.kpi.ua/node/179>.

Робочу програму навчальної дисципліни (силабус):

Складено: доцент, к.філол.н, Івашкевич Леся Сергіївна

Ухвалено: кафедрою ТППНМ (протокол № 6 від 15 січня 2021 р.)

Погоджено: Науково-методичною комісією університету зі спеціальності 035 Філологія
(протокол № 6 від 03.02.2021 р.)