

Силабус
освітнього компоненту ОК 16
(умовне позначення ОК в освітній програмі (ОП))

Системи машинного перекладу

Назва дисципліни:	Системи машинного перекладу
Рівень вищої освіти:	перший (бакалаврський)
Галузь знань:	В Культура, мистецтво та гуманітарні науки
Спеціальність:	В 11 «Філологія»
Спеціалізація:	В 11.10 Прикладна лінгвістика
Освітньо-професійна програма:	Філологія (Прикладна лінгвістика)
Сторінка курсу в Moodle:	https://dl2022.khadi-kh.com/course/view.php?id=5327
Рік навчання:	2
Семестр:	4
Обсяг освітнього компоненту	3 кредити (90 годин)
Форма підсумкового контролю	залік
Консультації:	за графіком
Назва кафедри:	Кафедра комп'ютерних наук і інформаційних систем
Мова викладання:	українська
Керівник курсу:	ст. викладач Неронов Сергій Миколайович
Контактний телефон:	+38-067-703-64-16
E-mail:	sernikner@gmail.com

Короткий зміст освітнього компоненту:

Метою є ознайомлення з теорією та практикою машинного перекладу. Студенти отримають розуміння основних підходів до машинного перекладу, вивчать ключові техніки та алгоритми, а також набудуть практичних навичок у розробці систем машинного перекладу. Курс також пропонує огляд сучасних тенденцій у галузі та вивчення найкращих практик машинного перекладу.

Предмет: вивчення теоретичних та практичних аспектів автоматизованого перекладу текстів різними мовами за допомогою комп'ютерних систем.

Основними завданнями вивчення навчальної дисципліни є:

- формування розуміння основних підходів до машинного перекладу: статистичного, нейромережевого та глибинного навчання;
- формування знань про різні види машинного перекладу, включаючи переклад на основі правил, статистичний переклад та нейромережевий переклад;
- формування розуміння принципів побудови моделей машинного перекладу, включаючи модель мови, модель перекладу та механізми уваги;
- формування розуміння оцінки якості машинного перекладу та використання метрик, таких як BLEU, METEOR та інші.

По завершенні вивчення дисципліни студенти повинні вміти:

- розробляти та тренувати статистичні та нейромережеві моделі перекладу;
- використання інструментів та бібліотек для розробки систем машинного перекладу, таких як, PyTorch, OpenNMT тощо;

- використовувати основні прийоми і програмні засоби для аналізу та дизасемблювання програмних продуктів з метою їх подальшого несанкціонованого використання,
- мати навички роботи з системами машинного перекладу, включаючи розробку та оцінку моделей, а також розуміння найкращих практик у цій галузі.

Ці знання та навички дозволять студенту розуміти принципи та методи машинного перекладу, а також здати їх у практичну дію при розробці та вдосконаленні систем машинного перекладу.

Передумови для вивчення освітнього компоненту:

для успішного оволодіння дисципліною студенти повинні оперувати знаннями, навичками та вміннями, отриманими у ході вивчення курсів ОК 8 Комп'ютерна графіка, ОК 9 Системи управління базами даних, ОК 20 Навчальна практика

Компетентності, яких набуває здобувач:

Загальні компетентності:

- ЗК 5. Здатність учитися й оволодівати сучасними знаннями.
- ЗК 11. Здатність застосовувати знання у практичних ситуаціях.
- ЗК 12. Навички використання інформаційних і комунікаційних технологій.

Спеціальні (фахові) компетентності:

СК 8. Здатність вільно оперувати спеціальною термінологією для розв'язання професійних завдань.

СК 12. Здатність до організації ділової комунікації.

СК 13. Здатність використовувати базові знання з інформаційних технологій для ефективного виконання професійних завдань.

Результати навчання відповідно до освітньої програми:

ПРН 2. Ефективно працювати з інформацією: добирати необхідну інформацію з різних джерел, зокрема з фахової літератури та електронних баз, критично аналізувати й інтерпретувати її, впорядковувати, класифікувати і систематизувати.

ПРН 6. Використовувати інформаційні і комунікаційні технології для вирішення складних спеціалізованих завдань і проблем професійної діяльності.

ПРН 20. Застосовувати професійно-комунікаційні навички для забезпечення ефективної діяльності в галузі інформаційних технологій.

ПРН 21. Використовувати організаційні та аналітичні навички для управління інформаційними процесами та оптимізації роботи в цифровій технологічній сфері.

Тематичний план

№ теми	Назва тем (ЛК, ЛР, ПР, СЗ, СР)	Кількість годин
		очна
1	ЛК Визначення та основні концепції машинного перекладу.	2
	ПР Історія розвитку машинного перекладу	2
	СР Види машинного перекладу	7
2	ЛК Автоматичний переклад на основі правил. Правила та граматики в машинному перекладі	2
	ПР Системи перекладу на основі правил	2
	СР Оцінка та покращення систем перекладу на основі правил	7
3	ЛК Статистичний машинний переклад	2

	ПР (ЛР, СЗ) Модель мови та модель перекладу	2
	Підходи до побудови статистичних моделей перекладу	
	СР Оцінка та покращення статистичних систем перекладу	7
4	ЛК Нейромережевий машинний переклад	2
	ПР (ЛР, СЗ) Популярні архітектури нейромережевих систем перекладу (PyTorch, OpenNMT, Seq2Seq, Transformer)	2
	СР Тренування та оцінка нейромережевих систем перекладу (наприклад, використання автокодувачів, механізми уваги)	7
5	ЛК Глибинне навчання в машинному перекладі	2
	ПР (ЛР, СЗ) Використання предобробки та переднавчання	2
	Застосування предобробки в машинному перекладі	
	СР Архітектури глибинного навчання для машинного перекладу	7
6	ЛК Оцінка та оцінювання якості машинного перекладу	2
	ПР (ЛР, СЗ) Метрики оцінки якості перекладу (наприклад, BLEU, METEOR)	2
	СР Людська оцінка якості перекладу	7
7	ЛК Сучасні тенденції та застосування машинного перекладу	2
	ПР (ЛР, СЗ) Машинний переклад у веб-додатках та сервісах	2
	Машинний переклад для специфічних галузей (медицина, юриспруденція, технічний переклад і т.д.)	
	СР Машинний переклад для мов, які мають обмежений обсяг ресурсів	8
8	ЛК Найкращі практики та майбутні напрямки	2
	ПР (ЛР, СЗ) Перспективи розвитку машинного перекладу	2
	СР Етичні питання в машинному перекладі	8
Разом	ЛК	16
	ПР (ЛР, СЗ)	16
	СР	58

Індивідуальне навчально-дослідне завдання:

Детальний розгляд студентами окремих теоретичних положень навчальної дисципліни і формування вміння та навичок їх практичного застосування шляхом виконання поставлених задач.

Методи навчання:

МН1– комунікативний метод (здійснення інтерактивної діяльності, навчальна дискусія, пояснення);

МН2 – практичний метод (виконання практичних завдань, написання наукових робіт);

МН3 – аудіо-візуальний метод (зорова невербальна наочність, сприйняття мови на слух: ілюстрації, аудіо, відео).

МН4 – метод спостереження (цілеспрямоване сприйняття мовних та мовленнєвих явищ, їх імітація та практична реалізація);

МН5 – самостійна робота як метод навчання (самостійне опрацювання теоретичного матеріалу з метою його практичного використання).

Форми та методи оцінювання

Опитування, перевірка виконання вправ і завдань, написання самостійних робіт, презентації результатів виконаних завдань та досліджень, поточне тестування, модульне тестування, залік.

Система оцінювання та вимоги:

Поточна успішність

Поточна успішність здобувачів за виконання навчальних видів робіт на навчальних заняттях і за виконання завдань самостійної роботи оцінюється за допомогою чотирибальної шкали оцінок з наступним перерахуванням у 100-бальну шкалу. Під час оцінювання поточної успішності враховуються всі види робіт, передбачені навчальною програмою.

Поточний контроль реалізується у формі опитування, експрес-контролю, контролю засвоєння навчального матеріалу, правильності виконання практичних завдань та тестів, у тому числі запланованих на самостійне опрацювання.

Здобувач освіти накопичує бали протягом семестру за виконання завдань, виступи на заняттях, участь у дискусіях, виступ на конференції, а також за виконання контрольних робіт.

Лекційні заняття оцінюються шляхом визначення якості виконання конкретизованих завдань.

Практичні заняття оцінюються якістю виконання контрольного або індивідуального завдання, виконання та оформлення практичної роботи.

Оцінювання поточної успішності здобувачів вищої освіти здійснюється на кожному практичному занятті за чотирибальною шкалою («5», «4», «3», «2») і заносяться у журнал обліку академічної успішності.

– «відмінно»: здобувач бездоганно засвоїв теоретичний матеріал, демонструє глибокі знання з відповідної теми або навчальної дисципліни, основні положення;

– «добре»: здобувач добре засвоїв теоретичний матеріал, володіє основними аспектами з першоджерел та рекомендованої літератури, аргументовано викладає його; має практичні навички, висловлює свої міркування з приводу тих чи інших проблем, але припускається певних неточностей і похибок у логіці викладу теоретичного змісту або при аналізі практичного;

– «задовільно»: здобувач в основному опанував теоретичні знання навчальної теми або дисципліни, орієнтується у першоджерелах та рекомендованій літературі, але непереконливо відповідає, плутає поняття, невпевнено відповідає на додаткові питання, не має стабільних знань; відповідаючи на питання практичного характеру, виявляє неточність у знаннях, не вміє оцінювати факти та явища, пов'язувати їх із майбутньою професією;

– «незадовільно»: здобувач не опанував навчальний матеріал теми (дисципліни), не знає наукових фактів, визначень, майже не орієнтується в першоджерелах та рекомендованій літературі, практичні навички не сформовані.

Підсумковий бал за поточну діяльність визнається як середньоарифметична сума балів за кожне заняття, за індивідуальну роботу, поточні контрольні роботи за формуюлю:

$$K^{поточ} = \frac{K1 + K2 + \dots + Kn}{n},$$

де $K^{поточ}$ – підсумкова оцінка успішності за результатами поточного контролю;

$K1, K2, \dots, Kn$ – оцінка успішності n -го заходу поточного контролю;

n – кількість заходів поточного контролю.

Оцінки конвертуються у бали згідно шкали перерахунку (таблиця 1).

Таблиця 1 – Перерахунок середньої оцінки за поточну діяльність у багатобальну шкалу

4-бальна шкала	100-бальна шкала	4-бальна шкала	100-бальна шкала	4-бальна шкала	100-бальна шкала	4-бальна шкала	100-бальна шкала
5	100	4,45	89	3,90	78	3,35	67
4,95	99	4,4	88	3,85	77	3,3	66
4,9	98	4,35	87	3,80	76	3,25	65
4,85	97	4,3	86	3,75	75	3,2	64
4,8	96	4,25	85	3,7	74	3,15	63
4,75	95	4,20	84	3,65	73	3,1	62
4,7	94	4,15	83	3,60	72	3,05	61
4,65	93	4,10	82	3,55	71	3	60
4,6	92	4,05	81	3,5	70	від 1,78 до 2,99	від 35 до 59
						повторне складання	
4,55	91	4,00	80	3,45	69	від 0 до 1,77	від 0 до 34
4,5	90	3,95	79	3,4	68	повторне вивчення	

Підсумкове оцінювання

- Здобувач вищої освіти отримує залік на останньому занятті з дисципліни за результатами поточного оцінювання. Середня оцінка за поточну діяльність конвертується у бали за 100-бальною шкалою, відповідно до таблиці перерахунку (таблиця 1).

Здобувачі вищої освіти, які мають середню поточну оцінку з дисципліни нижче ніж «3» (60 балів), на останньому занятті можуть підвищити свій поточний бал шляхом складання тестів з дисципліни.

Оцінювання знань здобувачів шляхом тестування здійснюється за шкалою:

- «Відмінно»: не менше 90 % правильних відповідей;
- «Дуже добре»: від 82 % до 89 % правильних відповідей;
- «Добре»: від 74 % до 81 % правильних відповідей;
- «Задовільно»: від 67 % до 73% правильних відповідей;
- «Задовільно достатньо»: від 60 % до 66 % правильних відповідей;
- «Незадовільно»: менше 60 % правильних відповідей.

- Умовою отримання заліку є:

- відпрацювання всіх пропущених занять;
- середня поточна оцінка з дисципліни не нижче «3» (60 балів).

- За виконання індивідуальної самостійної роботи та участь у наукових заходах здобувачам нараховуються додаткові бали.

3.1 Додаткові бали додаються до суми балів, набраних здобувачем вищої освіти за поточну навчальну діяльність (для дисциплін, підсумковою формою контролю для яких є залік), або до підсумкової оцінки з дисципліни, підсумковою формою контролю для якої є екзамен.

3.2 Кількість додаткових балів, яка нараховується за різні види індивідуальних завдань, залежить від їх об'єму та значимості:

- призові місця з дисципліни на міжнародному / всеукраїнському конкурсі наукових студентських робіт – 20 балів;
- призові місця з дисципліни на всеукраїнських олімпіадах – 20 балів;
- участь у міжнародному / всеукраїнському конкурсі наукових студентських робіт – 15 балів

– участь у міжнародних / всеукраїнських наукових конференціях студентів тамолодих вчених – 12 балів;

– участь у всеукраїнських олімпіадах з дисципліни – 10 балів

– участь в олімпіадах і наукових конференціях ХНАДУ з дисципліни – 5 балів;

– виконання індивідуальних науково-дослідних (навчально-дослідних) завдань підвищеної складності – 5 балів.

3.3 Кількість додаткових балів не може перевищувати 20 балів.

4 Результат навчання оцінюється (*обрати потрібне*):

– за двобальною шкалою (зараховано/не зараховано) згідно з таблицею 2;

– за 100-бальною шкалою (для диференційованого заліку) згідно з таблицею 3. Підсумкова оцінка разом з додатковими балами не може перевищувати 100 балів.

Загальна підсумкова оцінка за вивчення навчальної дисципліни визначається згідно зі шкалою, наведеною в таблиці 2.

Таблиця 2 – Шкала оцінювання знань здобувачів за результатами підсумкового контролю з навчальної дисципліни

Оцінка в балах	Оцінка за національною шкалою		Оцінка за шкалою ЄКТС	
	екзамен	залік	Оцінка	Критерії
90-100	Відмінно	Зараховано	A	Теоретичний зміст курсу освоєний цілком, без прогалин, необхідні практичні навички роботи з освоєним матеріалом сформовані, усі передбачені програмою навчання навчальні завдання виконані, якість їхнього виконання оцінено числом балів, близьким до максимального
80–89			B	Теоретичний зміст курсу освоєний цілком, без прогалин, необхідні практичні навички роботи з освоєним матеріалом в основному сформовані, усі передбачені програмою навчання навчальні завдання виконані, якість виконання більшості з них оцінено числом балів, близьким до максимального
75-79	Добре	Зараховано	C	Теоретичний зміст курсу освоєний цілком, без прогалин, деякі практичні навички роботи з освоєним матеріалом сформовані недостатньо, усі передбачені програмою навчання навчальні завдання виконані, якість виконання жодного з них не оцінено мінімальним числом балів, деякі види завдань виконані з помилками

Оцінка в балах	Оцінка за національною шкалою		Оцінка за шкалою ЄКТС	
	екзамен	залік	Оцінка	Критерії
67-74	Задовільно		D	Теоретичний зміст курсу освоєний частково, але прогалини не носять істотного характеру, необхідні практичні навички роботи з освоєним матеріалом в основному сформовані, більшість передбачених програмою навчання навчальних завдань виконано, деякі з виконаних завдань, можливо, містять помилки
60-66			E	Теоретичний зміст курсу освоєний частково, деякі практичні навички роботи не сформовані, багато передбачених програмою навчання навчальних завдань не виконані, або якість виконання деяких з них оцінено числом балів, близьким до мінімального.
35-59	Незадовільно	Не зараховано	FX	Теоретичний зміст курсу освоєний частково, необхідні практичні навички роботи не сформовані, більшість передбачених програм навчання навчальних завдань не виконано, або якість їхнього виконання оцінено числом балів, близьким до мінімального; при додатковій самостійній роботі над матеріалом курсу можливе підвищення якості виконання навчальних завдань (з можливістю повторного складання)
0-34	Неприйнятно		F	Теоретичний зміст курсу не освоєно, необхідні практичні навички роботи не сформовані, усі виконані навчальні завдання містять грубі помилки, додаткова самостійна робота над матеріалом курсу не приведе до якого-небудь значущого підвищення якості виконання навчальних завдань (з обов'язковим повторним курсом)

Політика курсу:

- освоєння дисципліни передбачає обов'язкове відвідування практичних занять, а також самостійну роботу;
- самостійна робота передбачає вивчення окремих тем навчальної дисципліни, які винесені відповідно до програми на самостійне опрацювання, або ж були розглянуті стисло;
- всі завдання, передбачені програмою, мають бути виконані у встановлений термін;
- якщо здобувач вищої освіти відсутній на заняттях з поважної причини, він презентує виконані завдання під час самостійної підготовки та консультації викладача;
- під час вивчення курсу здобувачі вищої освіти повинні дотримуватись правил академічної доброчесності, викладених у таких документах: «Правила академічної доброчесності учасників освітнього процесу ХНАДУ»

(https://www.khadi.kharkov.ua/fileadmin/P_Standart/pologeniya/stvnz_67_01_dobroch_1.pdf), «Академічна доброчесність. Перевірка тексту академічних, наукових та кваліфікаційних робіт на плагіат» (https://www.khadi.kharkov.ua/fileadmin/P_Standart/pologeniya/stvnz_85_1_01.pdf), «Морально-етичний кодекс учасників освітнього процесу ХНАДУ» (https://www.khadi.kharkov.ua/fileadmin/P_Standart/pologeniya/stvnz_67_01_MEK_1.pdf).

Рекомендована література:

1. Ваховська О. В. Основи комп'ютерної лінгвістики: навчально-методичний посібник. Київ : КНЛУ, 2023. 112 с.
2. Артеменко Ю. О. Методичні вказівки щодо самостійної роботи з навчальної дисципліни «Комп'ютерні системи перекладу» для студентів спеціальності 035 Філологія. Кременчук, 2021.
3. Корольова Т.М., Жмаєва Н.С., Колчак Ю.І. Постредагування при машинному перекладі. Науковий вісник ПНПУ ім. К.Д. Ушинського. 2020. № 30. С. 102–119. URL: <https://www.lingstud.od.ua/archive/2020/30/7.pdf> (останнє звернення 25.01.2024)

Додаткові джерела:

1. Schaler, Reinhard Localization and Translation. Handbook of Translation Studies, 2010, URL: https://www.academia.edu/2124984/Localization_and_translation
2. Fonteyne, M., Tezcan, A., Macken, L. (2020). Literary machine translation under the magnifying glass: Assessing the quality of an NMT-translated detective novel on document level. The 12th International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA), 2023. P. 3783-3791.
3. Навігація робочим процесом перекладу: Вичерпний посібник (2023) <https://www.conveythis.com/uk/guide-to-translation-workflow-management-2023/>

Розробник
силабусу навчальної
дисципліни:



Сергій НЕРОНОВ

Гарант освітньо-
професійної
програми



Анастасія ПТУШКА

Завідувач кафедри
комп'ютерних наук і



Ганна ПЛЄХОВА

інформаційних систем