

Київський столичний університет імені Бориса Грінченка
Факультет української філології, культури і мистецтва
Кафедра української мови



“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної
та навчальної роботи

Олексій ЖИЛЬЦОВ

14 лютого 2024 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ЛІНГВІСТИКА КОМП'ЮТЕРНОГО АНАЛІЗУ І ОБРОБКИ ІНФОРМАЦІЇ:
МАШИННИЙ ПЕРЕКЛАД

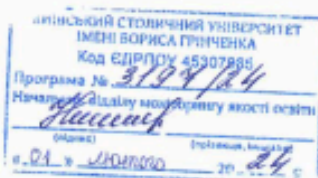
для студентів

рівень вищої освіти: перший (бакалаврський)

спеціальності: 035 Філологія

освітньої програми: 035.01.01 Українська мова і література

Київ – 2024



Розробник: Бойко Марія Іванівна, старший викладач кафедри української мови, доктор філософії.

Викладач: Бойко Марія Іванівна, старший викладач кафедри української мови, доктор філософії.

Робочу програму розглянуто і затверджено на засіданні кафедри української мови

Протокол від 17 січня 2024 р. № 1

Завідувач кафедри  (Михайло ВІНТОНІВ)

Робочу програму погоджено з гарантом освітньої програми

035.01.01 «Українська мова та література» першого (бакалаврського) рівня вищої освіти

31.01.2024 р.

Гарант освітньої програми  (Тетяна ВИДАЙЧУК)

Робочу програму перевірено

01.02.2024 р.

Заступник декана  (Світлана ГОРОБЕЦЬ)

Пролонговано:

на 20_/20_ н.р	_____ (підпис)	(_____) ПІБ	«__»__ 20 __ р.	протокол№ __
на 20_/20_ н.р	_____ (підпис)	(_____) ПІБ		протокол№__

			«__»__ 20 __ р.	
--	--	--	--------------------	--

1. Опис навчальної дисципліни «Лінгвістика комп'ютерного аналізу і обробки інформації»

Найменування показників	Характеристика дисципліни за формами навчання			
	денна	заочна		
Вид дисципліни	вибіркова			
Мова викладання, навчання та оцінювання	українська			
Загальний обсяг кредитів / годин	20/ 600			
Курс	2-4			
Семестр	3-7			
Кількість змістових модулів з розподілом:	5			
<i>Змістовий модуль: Машинний переклад</i>				
Семестр	6			
Обсяг кредитів	4			
Обсяг годин, у тому числі:	120			
Аудиторні	56			
Модульний контроль	8			
Семестровий контроль	-			
Самостійна робота	56			
Форма семестрового контролю	залік			

2. Мета та завдання навчальної дисципліни

Мета навчальної дисципліни: формування інформаційної компетентності у студентів; знань про специфіку дисципліни «Лінгвістика комп'ютерного аналізу й обробки інформації: Машинний переклад», історію, особливості і сфери застосування машинного перекладу, алгоритми роботи систем МП, можливості використання автоматичного і автоматизованого машинного перекладу; набуття студентами комплексу практичних навичок і умінь, необхідних для роботи з програмними продуктами машинного перекладу.

Завдання навчальної дисципліни: формування у студентів знань про історію і становлення машинного перекладу (МП) як наукового напрямку, основні проблеми і завдання, які вирішуються за допомогою МП, теоретичні і практичні результати взаємодії і взаємовпливу інформатики та лінгвістики; опанування базовими базовими поняттями машинного перекладу; набуття вмінь навичок роботи з програмами програм на зразок "пам'ять перекладача" і програм машинного перекладу; оволодіння прийомами ефективного пошуку необхідної термінології з використанням словників, глосаріїв, баз даних, в тому числі в Інтернеті, результативного використання можливостей машинного та автоматизованого перекладу для вирішення перекладацьких завдань; формування умінь роботи з електронними словниками та іншими електронними ресурсами для розв'язання лінгвістичних задач; набуття навичок користування основними інформаційно-пошуковими і експертними системами, системами подання знань, синтаксичного і морфологічного аналізу, автоматичного синтезу і розпізнавання мови, обробки лексикографічної інформації та автоматизованого перекладу.

Під час практичних занять, індивідуальної навчально-дослідницької та самостійної роботи студенти, майбутні спеціалісти **набувають таких програмних компетентностей: ЗК (загальні компетентності), ФК(фахові компетентності), ПР-3 (програмні результати).**

3. Результати навчання за дисципліною

ЗК–3 Інформаційна компетенція. Здатність до самостійного пошуку та оброблення інформації з різних джерел для розгляду конкретних питань; до ефективного використання інформаційних технологій в соціальній і професійній діяльності. Сервіс-орієнтування.

ЗК–7 Самоосвітня компетенція. Здатність до самостійної пізнавальної діяльності, самоорганізації та саморозвитку. Спрямованість на розкриття особистісного потенціалу та самореалізацію. Прагнення до особистісно-професійного лідерства та успіху.

ФК–1 Філологічна (вільне володіння базовим категорійнопоняттєвим та аналітично-дослідницьким апаратом сучасної філології) -теоретико

методологічна та категоріальна обізнаність у загальнонаукових та спеціальних методах та дослідницьких підходах розуміння естетико-практичної значущості філологічного знання; - володіння ключовими поняттями мовознавства, літературознавства та фольклористики, теоріями і методами філологічного аналізу, особливостями розвитку філологічної думки у певну історичну епоху та у сучасному контексті розвитку гуманітаристики.

ФК-2 Лінгвістична (глибокі знання та розуміння в галузі лінгвістики).

Здатність аналізувати лінгвістичні явища з погляду фундаментальних філологічних принципів і знань, а також на основі відповідних загальнонаукових методів. Знання природи і сутності мови, походження й закономірностей розвитку мов, особливостей структури мови, класифікації мов світу, шляхів і методів вивчення мовного матеріалу. Знання системи сучасної української літературної мови, еволюції її граматичної будови.

ФК-5 Прикладна лінгвістично-орієнтована. Здатність застосовувати нелінгвістичні знання в практичній філологічній діяльності. Уміти застосовувати знання ІКТтехнологій для виконання лінгвістичних завдань. Володіти базовими знаннями з юриспруденції, вміти застосовувати їх під час аналізу мовної політики та мовної ситуації в державі, проведення експертиз юридично значимих текстів.

ПРН-3-6 Володіння методами і прийомами здійснення усної та писемної комунікації; методами і прийомами створення різних типів і видів текстів (мовних творів); методами і прийомами аналізу й інтерпретації різних типів і видів текстів (мовних творів); методами і прийомами обробки та трансформації різних типів і видів текстів (творів).

ПРН-3-7 Знання основних бібліографічних покажчиків і пошукових систем, володіння прийомами бібліографічного опису різних джерел; сучасними методами і прийомами обробки інформації в галузі професійної діяльності.

ПРН-3-9 Здатність застосовувати нелінгвістичні знання у філологічній діяльності. Знання мовного законодавства, уміння застосовувати його у філологічній та педагогічній діяльності, аналізувати конфліктні та юридично значимі тексти. Оволодіння навичками застосовувати ІКТ- технології, лінгвістичні програмні продукти для створення баз даних, автоматизованих лексикографічних систем, систем машинного перекладу і користування ними.

ПРН-У-1 Вільне володіння базовим професійним категоріальнопоняттєвим та дослідницьким апаратом сучасної філологічної науки. Уміння пояснювати професійно з історичних позицій явища і факти української мови та літератури.

ПРН-У-7 Ефективно застосовувати ІКТ-технології у науково-дослідній і науково-педагогічній діяльності.

ПРН-У-8 Уміння працювати самостійно, працювати з метою підвищення фахової компетентності, самоосвіти; уміння працювати в умовах обмеженого часу з акцентом на професійну сумлінність та унеможливлення плагіату.

ПРН-С-1 Фахово формувати й аргументувати власні теоретичні і практичні судження, професійні позиції.

ПРН-С-3 Здатність формувати судження про значення і наслідки своєї професійної діяльності з урахуванням соціальних, професійних і етичних позицій

Організація навчальної діяльності в Центрі культури української мови

Реформування системи вищої освіти на сучасному етапі зумовлює концептуальні зміни в напрямках, цілях, завданнях, змісті освіти, філологічної зокрема, які зорієнтовують передовсім на розвиток особистості, формування творчої ініціативи, мобільності, підвищення конкурентоздатності майбутнього фахівця.

Задля ефективної реалізації концепції нової освітньої стратегії університету програмою передбачено застосування можливостей Центру культури української мови.

Для цього впроваджено систематичне застосування інноваційних тренінгів, дослідницьких завдань, що передбачають карткування укладання мультимедійних словників, опрацювання спеціальної літератури тощо.

Крім того, студентів залучено до роботи в проблемних групах, до участі в наукових конференціях; мовно-інформаційних конкурсах тощо.

У Центрі заплановано презентації наукових видань з проблем культури української мови, зустрічі з науковцями.

4. Структура навчальної дисципліни

Назви теоретичних розділів	Кількість відведених годин				Підсумковий контроль
	Лекції	Семінарські заняття	Практичні заняття	Самостійна робота	
Змістовий модуль 1. Машинний переклад як напрям штучного інтелекту.					
Тема 1. Машинний переклад як напрямок штучного інтелекту	2			2	
Тема 2. Види і стратегії машинного перекладу	2			2	
Тема 3. Структура систем машинного перекладу.	2			2	
Тема 4. Лінгвістичні проблеми машинного перекладу	2			2	
Тема 5. Використання паралельних корпусів текстів в машинному перекладі. Пам'ять перекладів	2			2	
Тема 6. Основні концептуальні ідеї і принципи машинного перекладу.		2		2	
Тема 7. Можливості мережі Інтернет у роботі перекладача. Методики ефективного пошуку лінгвістичної інформації у мережі Інтернет.		2		2	
Модульний контроль №1					2
Змістовий модуль 2. Становлення та розвиток машинного перекладу. Машинний переклад і комп'ютерна лінгвістика.					
Тема 8. Історія машинного перекладу.		2		2	
Тема 9. Зв'язок машинного перекладу з іншими областями знань		2		2	

Тема 10. Комп'ютер на місці перекладача.		2		2	
Тема 11. Компанія GOOGLE та її перекладацькі сервіси.		2		2	
Модульний контроль № 2					2
Змістовий модуль 3. Автоматичні перекладні словники і електронні системи управління термінологією					
Тема 12. Автоматичні перекладні словники і електронні системи управління термінологією.		2		2	
Тема 13. Двомовні словники в складі систем машинного перекладу.		2		2	
Тема 14. Лінгвістичні основи інформаційних технологій в перекладі.		2		2	
Тема 15. Лінгвістичні основи інформаційних технологій в перекладі		2		2	
Тема 16. Попереднє редагування і постредування тексту.		2		2	
Тема 17. Проблеми машинного перекладу – типологія помилок.		2		2	
Тема 18. Типи систем машинного перекладу.		2		2	
Тема 19. Типи систем машинного перекладу.		2		2	
Тема 20. Алгоритм машинного перекладу, заснованого на лінгвістичному аналізі.		2		2	
Модульний контроль №3					2
Змістовий модуль 4. Пам'ять перекладів як явище.					

Тема 21. Аналіз алгоритмів роботи систем машинного перекладу.		2		2	
Тема 22. Пам'ять перекладу.		2		2	
Тема 23. Лінгвістичні проблеми машинного перекладу.		2		2	
Тема 24. Підготовка до роботи у середовищі системи автоматизованого перекладу з підтримкою пам'яті перекладів.		2		2	
Тема 25. Робота з системою автоматизованого перекладу OmegaT.			2	2	
Тема 26. Робота з укладанням та використанням глосарію OmegaT			2	2	
Тема 27. Робота з укладанням та використанням глосарію OmegaT			2	2	
Тема 28. Виконання галузевого перекладу на базі OmegaT			2	2	
Модульний контроль № 4					2
Підсумковий контроль – залік					
Усього годин: 120	10	38	8	56	8

5. Програма навчальної дисципліни

ЗМІСТОВИЙ МОДУЛЬ 1.

Історія розвитку машинного перекладу. Огляд наявних систем автоматичного і автоматизованого перекладу.

Теми лекцій

Лекція 1

Машинний переклад як напрямок штучного інтелекту. Визначення машинного перекладу. Короткі відомості з історії розвитку напрямку. Стимули до розвитку досліджень. Переваги машинного перекладу.

Лекція 2

Види і стратегії машинного перекладу. Класифікація перекладів. Еволюція машинного перекладу. Типи машинного перекладу за ступенем автоматизації. Стратегії машинного перекладу.

Лекція 3

Структура систем машинного перекладу. Допоміжні програмні засоби. Склад лінгвістичної бази даних. Лінгвістичний процесор.

Лекція 4

Лінгвістичні проблеми машинного перекладу. Проблема багатозначності при машинному перекладі. Синтаксичні трансформації в машинному перекладі. Переклад фразеологічних поєднань в системах машинного перекладу.

Лекція 5

Використання паралельних корпусів текстів в машинному перекладі. Пам'ять перекладів. Види і структура паралельних корпусів текстів. Технологія пам'яті перекладів. Переваги та недоліки технології.

Теми семінарських занять

Семінарське заняття 1

Основні концептуальні ідеї та принципи машинного перекладу.

План заняття:

1. Предмет, цілі, завдання та зміст курсу.
2. Визначення машинного перекладу.
3. Предмет і об'єкт наукового напрямку.
4. Доцільність створення систем МП. Застосування систем МП. Переваги МП.
5. Проблеми інформатизації філологічної освіти.
6. Машинний переклад і бізнес.

Завдання:

1. Підібрати приклади використання машинного перекладу представниками різних професій, обґрунтувати його необхідність.
2. Оцінити перспективи машинного перекладу, знайти приклади повністю автоматичного перекладу в науково-фантастичній літературі, обґрунтувати реальність (нереальність) існування таких систем в майбутньому.

Семінарське заняття 2

Можливості мережі Інтернет у роботі перекладача. Методики ефективного пошуку лінгвістичної інформації у мережі Інтернет.

План заняття:

1. Інтернет як глобальна інформаційне середовище. Мовні моделі в Інтернет.
2. Інформаційно-пошукові системи. Аналіз особливостей роботи пошукових систем.

3. Енциклопедії та довідкові матеріали у вільному доступі в мережі.
4. Лінгвістичний пошук в Інтернет.
5. Тематичний пошук в мережі.
6. Загальні рекомендації по роботі з пошуковими системами. Поради з пошуку в Інтернеті.
7. Перекладацькі та лінгвістичні сайти.

Завдання:

1. Здійснити пошук інформації засобами 3-4 різних пошукових систем. Порівняти результати пошуку та їх релевантність. Оформити звіт про виконання.

Питання для обговорення:

1. Розкажіть про основні сфери діяльності, в яких інтернет незамінний для перекладача?
2. Розкажіть про пошук лінгвістичної інформації в інтернеті.
3. Що цікавого для перекладача можна знайти на перекладацьких і лінгвістичних сайтах?
4. Опишіть два основних стимули розвитку робіт з машинного перекладу в сучасному світі.

ЗМІСТОВИЙ МОДУЛЬ 2.

Становлення та розвиток машинного перекладу. Машинний переклад і комп'ютерна лінгвістика.

Семінарське заняття 3

Історія машинного перекладу.

План заняття:

1. Становлення і розвиток прикладного (комп'ютерного) напрямку в лінгвістиці. Перші спроби машинного перекладу.
2. 1940-і роки - перші розробки в області створення систем МП.
3. 1950-ті роки - труднощі в створенні систем МП, їх причини.
4. 1960-і роки - створення експериментальних лабораторій МП.
5. 1970-80-ті роки - розвиток обчислювальної техніки як нового потужного імпульсу в роботі по створенню перекладацьких систем.
6. 1990-ті роки - бурхливий розвиток ринку ПК, поширення мережі Інтернет.
7. Сучасна історія машинного перекладу.

Завдання:

1. Підготувати усні повідомлення з машинного перекладу першого, другого, третього поколінь.

Питання для обговорення:

1. Коли з'явилися перші програми МП і коли вони набули широкого поширення?
2. Назвіть одиниці перекладу.
3. Чим охарактеризовані 1940-і роки в розвитку МП?
4. Чим охарактеризовані 1950-і роки в розвитку МП?
5. Чим охарактеризовані 1970-80-ті роки у розвитку МП?

6. Чим охарактеризовані 1990-і роки в розвитку МП?

Семінарське заняття 4

Зв'язок машинного перекладу з іншими областями знань

План заняття:

1. Машинний переклад в контексті досліджень в області штучного інтелекту.
2. Комп'ютерна лінгвістика на сучасному етапі. Виникнення нових інформаційних областей.
3. Машинний переклад і комп'ютерна лінгвістика.
4. Зв'язок машинного перекладу з математичною лінгвістикою і психолінгвістикою.

Завдання:

1. Провести дослідження наукової літератури із зазначеної тематики, підібрати приклади і підготувати доповідь на одне з питань.

Семінарське заняття 5

Комп'ютер на місці перекладача.

План заняття:

1. Три типи систем машинного перекладу за ступенем участі людини.
2. Інструменти для перекладу в Інтернеті. Електронні словники, глосарії, енциклопедії та довідники.
3. Онлайн перекладачі. Словники онлайн.
4. Послуги з перекладу в Інтернет.
5. Комп'ютерні програми підтримки словників (бази даних, електронні картотеки, програми обробки текстів).

Завдання:

1. Проаналізувати можливості електронного словника, встановленого на пристрої, з яким працює студент, а також онлайн-словника. Здійснити порівняльну характеристику. Оформити звіт про виконання.

Питання для обговорення:

1. Розкажіть про способи взаємодії людини і комп'ютера при здійсненні МП.
2. Чи може комп'ютер замінити перекладача?
3. Охарактеризуйте відмінності он-лайн і оф-лайн сервісів.
4. Проаналізуйте технічні можливості словників.
5. Які службові функції може виконувати електронний словник?
6. Звукові можливості і особливості електронних словників.
6. Інтерактивні навчальні можливості словників.

Семінарське заняття 6
Компанія GOOGLE та її перекладацькі сервіси.

План заняття:

1. Компанія Google і її перекладацькі послуги.
3. Сервіси компанії Goggle.
4. Методика роботи Goggle перекладача. Google словник.

Питання для обговорення:

1. Мультимедійні та інтерактивні функції Goggle перекладача.
2. Назвіть додатки компанії Goggle, в яких присутні перекладацькі функції.
3. Мультимедійні та інтерактивні функції Goggle перекладача.
4. Назвіть додатки компанії Goggle, в яких присутні перекладацькі функції.
5. Розкажіть про функціональні можливості сервісу google traslate і google translation toolkit.

ЗМІСТОВИЙ МОДУЛЬ 3.

Автоматичні перекладні словники і електронні системи управління термінологією.

Семінарське заняття 7

Автоматичні перекладні словники і електронні системи управління термінологією

План заняття:

1. Формування словника.
2. Становлення комп'ютерної лексикографії.
3. Основні принципи системності термінології.
4. Зовнішні та внутрішні ознаки типологізації терміносистеми.
5. Лексичний склад і обсяг автоматизованих перекладних словників.
6. Склад і структура словникових статей.
7. Динамічність автоматичних перекладних словників.
8. Місце перекладного словника в процесі перекладу.

Завдання:

1. Обрати одне з питань теми і підготувати доповідь. Інформацію оформити у вигляді презентації.
2. Проаналізувати якість машинного перекладу, виконаного засобами програмного продукту, встановленого на пристрої, а також за допомогою різних онлайн-сервісів. Порівняти якість перекладу в залежності від методу перекладу.
3. Побудувати три словникові статті українсько-англійського машинного словника за різними принципами (словник словоформ, словник квазіоснов). Вказати у словниках морфологічну, синтаксичну, семантичну інформацію про лексеми.

Семінарське заняття 8

Двомовні словники в складі систем машинного перекладу

План заняття:

1. Особливості двомовних машинних словників в порівнянні з одномовними.
2. Подання лексичних одиниць в двомовних машинних словниках.
3. Принципи відбору лексики для перекладних словників.
4. Кодування лінгвістичної інформації в словниках на морфологічному, синтаксичному, семантичному рівнях.
5. Комп'ютерна підтримка електронних бібліотек.

Завдання:

1. Підібрати 30 українськи лексем різних самостійних частин мови на основі частотних словників.
2. Виділити всі типи словозміни для обраних лексем.
3. Підібрати до них перекладні еквіваленти англійською мовою з урахуванням багатозначності.

Семінарське заняття 9

Лінгвістичні основи інформаційних технологій в перекладі

План заняття:

1. Переклад як аналітико-синтетичний процес. Операції, які забезпечують аналіз і синтез в системі машинного перекладу.
2. Моделі процесу перекладу.
3. Основні теоретичні моделі: ситуативна модель, багаторівнева модель, комплексна модель.
4. Основні принципи розв'язання лінгвістичних задач методом моделювання.

Завдання:

1. Об'єднатися у групи, вибрати одне з питань плану, підготувати повідомлення, інформацію оформити у вигляді презентації.

Семінарське заняття 10

Структура систем машинного перекладу

План заняття:

1. Структура лінгвістичної бази знань систем машинного перекладу.
2. Компоненти лінгвістичної бази даних.
3. Структура лінгвістичного процесора.

Завдання:

1. Розробити структуру лінгвістичної бази знань українсько-англійської системи машинного перекладу.
2. Виявити відмінності в структурі бази знань для українсько-англійської та англо-української систем машинного перекладу.

3. Виявити особливості морфологічного класифікатора для англійської мови в порівнянні з українським.

Семінарське заняття 11

Попереднє редагування і постредагування тексту

План заняття:

1. Поняття передредагування.
2. Основні напрями передредагування.
3. Лексико-семантичні і граматичні перетворення вхідного тексту у процесі передредагування.
4. Членування речень під час передредагування.
5. Зміна порядку слів у процесі передредагування.
6. Заміна частин мови і членів речення.
7. Поняття постредагування і його типів.
8. Особливості інформаційного і літературного постредагування.
9. Ситуативне постредагування.

Семінарське заняття 12

Проблеми машинного перекладу – типологія помилок.

Оцінка ефективності основних наявних систем машинного перекладу.

План заняття

1. Багатозначність і безкінечна різноманітність природньої мови на морфологічному, синтаксичному, лексичному і семантичному рівнях як причини недосконалості машинного перекладу.
2. Критерії оцінки якості машинного перекладу.
3. Шкала оцінки якості перекладу.
4. Якість перекладу на різних мовних рівнях.
5. Критерії вибору найбільш ефективних систем у практичній діяльності перекладацьких агенцій.

Завдання:

1. Познайомитися зі статтею Л.Маркеса «Automatic Evaluation of Machine Translation Quality»
[[Http://www.dialog-21.ru/digests/dialog2013/materials/pdf/1_MarquezL.pdf](http://www.dialog-21.ru/digests/dialog2013/materials/pdf/1_MarquezL.pdf)].
2. Використовуючи сервіс Asiya [<http://asiya.lsi.upc.edu/demo/>], оцінити якість перекладу декількох текстів, підібраних самостійно.
3. Описати сервіс Asiya, вказати його користь, новаторство, переваги і недоліки.
4. Розробити анкету для випробовуваних за певним текстом, використовуючи методику «вивчення відповідей за оригіналом тексту людини, читав тільки переклад », провести опитування, зробити висновки.

Семінарське заняття 13

Типи систем машинного перекладу

План заняття:

1. Системи прямого (послідовного) перекладу.
2. Переклад з допомогою синтаксичного переносу.
3. Статистичний машинний переклад. Проблеми адекватності машинного перекладу, виконаного методом статистики. Інтерактивність методу статистики в автоматизованому перекладі.
4. Важливість (сутність) перекладача в процесі (редагування машинного автоматичного) автоматизованого перекладу.

Завдання:

1. Виконати переклад галузевого тексту за допомогою програмного забезпечення, яке працює за методом статистики. Здійснити редагування отриманого тексту.
2. Запропонувати 5-10 власних варіантів перекладу і внести їх у базу даних програмного забезпечення.
3. Оформити звіт про виконання.

Семінарське заняття 14

Типи систем машинного перекладу

План заняття:

1. Редагування машинного перекладу, виконаного методом мовних правил.
2. Трансферні системи.
3. Системи семантичного типу.
4. Технології Transfer і Interlingua в системах машинного перекладу.
5. Три типи систем машинного перекладу за ступенем участі людини.

Питання для обговорення:

1. Розкажіть про способи взаємодії людини і комп'ютера при здійсненні МП.
2. Охарактеризуйте три типи систем МП FАMT, НАMT, МАNT.
3. Чи може комп'ютер замінити перекладача?
4. На чому заснована концепція МП Interlingva?
5. На чому заснована концепція МП Transfer?
6. Розкажіть про відмінності систем МП заснованих на технології Interlingua і Transfer.

Семінарське заняття 15

Алгоритм машинного перекладу, заснованого на лінгвістичному аналізі

План заняття:

1. Етапи морфологічного, синтаксичного, семантичного аналізу вихідного тексту.

2. Синтез тексту на перекладну мову: семантичне уявлення тексту, побудова синтаксичних структур, морфологічний синтез.

Завдання:

1. Перевести текст за допомогою різних систем машинного перекладу.
2. Виявити складності, з якими стикаються системи.
3. Провести морфологічний аналіз вихідного тексту, побудувати дерева залежностей речень, встановити семантичні зв'язки всередині кожного з пропозицій.
4. Назвіть етапи синтезу перекладного тексту, на яких виявлені помилки можуть бути виправлені.
5. Розробити рекомендації щодо поліпшення якості роботи систем машинного перекладу.

ЗМІСТОВИЙ МОДУЛЬ 4.

Пам'ять перекладів як явище.

Семінарське заняття 16

Аналіз алгоритмів роботи систем машинного перекладу.

План заняття:

1. Основні складові системи МП: словники, лексичні одиниці, правила.
2. Шляхи вирішення проблеми комунікації: створення діалогових систем і систем "обробки 5 природної мови".
3. Граматика в системах машинного перекладу.
4. Різновиди перекладу в залежності від жанрового типу перекладного матеріалу.
5. Лексико-граматичні особливості науково-технічних матеріалів.
6. Машинний переклад художніх текстів.

Завдання:

1. Виконати переклад художнього тексту за допомогою програмного забезпечення, яке працює за методом мовних правил. Здійснити редагування отриманого тексту.
2. Запропонувати 5-10 власних варіантів перекладу і внести їх у базу даних програмного забезпечення.
3. Оформити звіт про виконання.

Семінарське заняття 17

Пам'ять перекладу.

План заняття:

1. Відмінність програмного забезпечення з підтримкою пам'яті перекладів від інших типів програм автоматизованого перекладу.
2. Види перекладацьких програм з підтримкою пам'яті перекладів.

3. Класифікація програмних продуктів з підтримкою пам'яті перекладів.
4. Можливості програм пам'яті перекладу.
5. Сумісність і взаємозамінність програм.
6. Програми автоматичної перевірки якості перекладу.

Семінарське заняття 18

Лінгвістичні проблеми машинного перекладу

План заняття:

1. Проблеми машинного перекладу на морфологічному і синтаксичному рівнях.
2. Формальні і змістовні невідповідності близькоспоріднених і неблізкородственних мов.
3. Проблеми машинного перекладу на семантичному і прагматичному рівнях.
4. Облік контексту і екстралінгвістичних чинників при вирішенні полісемії.
5. Варіанти вирішення семантичних складнощів і підвищення якості машинного перекладу.

Завдання:

1. Вибрати 5 російських і 5 англійських уривків текстів різних стилів: науково-технічного, публіцистичного, художнього, розмовного, віршованого.
2. Перевести з допомогою п'яти різних систем машинного перекладу.
3. Виявити помилки в перекладах на різних мовних рівнях, запропонувати шляхи їх вирішення.
4. Знайти в початкових текстах омоніми (слова, що мають кілька перекладних еквівалентів), скласти для них статті контекстологічного словника.

Семінарське заняття 19

Підготовка до роботи у середовищі системи автоматизованого перекладу з підтримкою пам'яті перекладів.

1. Система перекладів OmegaT.
2. Системні вимоги і принцип роботи програми OmegaT.
3. Підтримувані формати файлів.
4. Компоненти системи OmegaT.
5. Глосарій як інструмент перекладача.
6. Дані, необхідні для укладання глосарію.
7. Методи укладання глосарію.
8. Структура глосарної статті.
9. Інтегрування даних глосарію з пам'яттю перекладів.
10. Використання глосарію для редагування виконаних перекладів.

Теми практичних занять

Практичне заняття 1

Робота з системою автоматизованого перекладу OmegaT

Завдання:

1. Ознайомитися з можливостями системи автоматизованого перекладу OmegaT.
2. Створити проект перекладу та власний глосарій.
3. Оформити звіт про виконання.

Практичне заняття 2

Робота з укладанням та використанням глосарію OmegaT

Завдання:

1. Ознайомитися з структурою пам'яті перекладів OmegaT.
2. Створити папку глосарію та додати терміни з поточного проекту.
3. Оформити звіт про виконання.

Практичне заняття 3

Виконання галузевого перекладу на базі OmegaT

Завдання:

1. Виконати переклад галузевого тексту з використанням глосарію перекладених раніше шаблонів.
2. Порівняти обсяг тексту оригіналу та перекладу.
3. Обчислити швидкість перекладу для конкретного типу тексту та напрямку перекладу.
4. Оформити звіт про виконання.

Практичне заняття 4

Виконання галузевого перекладу з використанням глосарію OmegaT

Завдання:

1. Виконати галузевий переклад з використанням глосарію перекладених раніше шаблонів.
2. Порівняти обсяг тексту оригіналу та перекладу.
3. Обчислити швидкість перекладу, враховуючи використання

6. Система оцінювання навчальних досягнень студентів

Вид діяльності студента	Макс. к.бали в за од.	Модуль 1		Модуль 2		Модуль 3		Модуль 4	
		К.од.	М.к.б	К.од.	М.к.б	К.од.	М.к.б.	К.од.	М.к.б.
Відвідування лекцій	1	5	5						
Відвідування семінарських занять	1	2	2	4	4	9	9	4	4
Відвідування практичних занять	1							4	4
Робота на семінарському занятті	10	2	20	4	40	9	90	4	40
Робота на практичному занятті	10							4	40
Виконання завдання для самостійної роботи	5	3	15	3	15	3	15	3	15
Модульна контрольна робота	25	1	25	1	25	1	25	1	25
Разом			67		84		139		128
Максимальна кількість балів									418
Розрахунок коефіцієнта									4,18

Кожний модуль передбачає бали за поточну роботу студента на практичних заняттях, виконання самостійної роботи, модульну контрольну роботу.

Виконання модульних контрольних робіт здійснюється в режимі комп'ютерної діагностики або з використанням роздрукованих завдань.

Творчі роботи, які виконує студент за визначеною тематикою, обговорюються та захищаються на практичних заняттях.

Модульний контроль знань студентів здійснюється після завершення вивчення навчального матеріалу модуля.

Кількість балів за роботу з теоретичним матеріалом, на практичних заняттях, під час виконання самостійної та індивідуальної навчально-дослідної роботи залежить від дотримання таких вимог:

- ✓ своєчасність виконання навчальних завдань;
- ✓ повний обсяг їх виконання;
- ✓ якість виконання навчальних завдань;
- ✓ самостійність виконання;
- ✓ творчий підхід у виконанні завдань;
- ✓ ініціативність у навчальній діяльності.

6.2.Завдання для самостійної роботи та критерії її оцінювання КАРТА САМОСТІЙНОЇ РОБОТИ СТУДЕНТА

Змістовий модуль та теми курсу	Академічний рівень	Бали	Термін виконання (тижні)
Змістовий модуль 1. Машинний переклад як напрям штучного інтелекту.			
Тема 1. Підготувати презентацію однієї з наявних систем машинного перекладу і нових можливостей інтеграції систем МТ і ТМ	Практичне заняття, модульний контроль	15	
Змістовий модуль 2. Становлення та розвиток машинного перекладу. Машинний переклад і комп'ютерна лінгвістика.			
Тема 2. Скласти есе, що містить порівняльний аналіз перекладу, виконаного різними системами;	Практичне заняття, модульний контроль	15	
Змістовий модуль 3. Автоматичні перекладні словники і електронні системи управління термінологією			
Тема 3. Створити термінологічну базу в SDL Trados Multiterm по заданій тематиці;	Практичне заняття, модульний контроль	15	
Змістовий модуль 4. Пам'ять перекладів як явище.			
Тема 4. Виконати перекладацький проект (перевести текст, використовуючи програму на основі ТМ).	Практичне заняття, модульний контроль	15	

Критерії оцінювання виконання студентами завдань для самостійної роботи

Кількість балів	Оцінка
5	Представлений матеріал розкриває і обґрунтовує тему, завдання виконано в повному обсязі відповідно до етапів навчального перекладацького проєкту. Коректно і повно використано пам'ять перекладу на рівні фраз, термінів і словосполучень.
4	Представлений матеріал недостатньо обґрунтований, виконане завдання містить неточності, які не впливають на зміст виконаної роботи
3	Представлений матеріал недостатньо обґрунтований, виконане завдання містить неточності, які частково впливають на якість виконаної роботи
2	Є неточності у викладенні матеріалу, відсутня логічна послідовність виконаної роботи
1	Допущені фактичні помилки в змісті роботи

6.3. Форми проведення модульного контролю та критерії оцінювання

Кожний модуль включає бали за *модульну контрольну роботу* (модульний контроль), розміщену в електронному навчальному курсі. Модульний контроль здійснюється після завершення вивчення навчального матеріалу модуля.

Критерії оцінювання виконання студентами модульних контрольних робіт

21-25 балів – завдання модульної контрольної роботи виконані в повному обсязі, лаконічно, змістовно, обґрунтовано, із наведенням прикладів і застосуванням методу аналізу та творчого підходу. Студент самостійно працює оцінює комунікативні явища, професійні ситуації; робить висновки; вміє формулювати й обґрунтовувати власну позицію; креативно підходить до їх підготовки; має практичні вміння і навички.

15-20 балів – завдання модульної контрольної роботи виконані в повному обсязі, лаконічно, змістовно, обґрунтовано, але з певними помилками щодо викладу матеріалу та без наведення прикладів до окремих фактів і явищ.

Студент робить висновки; Однак у процесі відтворення самостійно дібраного матеріалу простежується брак власних суджень, прояву креативності й ініціативності.

9-14 балів – завдання модульної контрольної роботи виконані в повному обсязі, але з порушенням логіки й послідовності викладу матеріалу та без ілюстрування прикладами. У роботі допускаються стилістичні помилки, неточне вживання термінів, довільне витлумачення фактів. Студент не висвітлює власної позиції; не втілює знання у практику допускаючи грубі помилки та уникаючи використання фахової термінології.

3-8 балів – завдання модульної контрольної роботи виконані не в повному обсязі, частково самостійно і за певним зразком. Студент володіє матеріалом на початковому рівні, викладає його хаотично і необґрунтовано; має фрагментарні навички роботи з науковими джерелами і не вміє робити висновки; не втілює знання у практику; практичні вміння і навички мають низький рівень розвитку.

0-2 бали – виконано менше половини завдань для самостійної роботи. Матеріал фрагментарний, не систематизований, не ілюстрований прикладами, порушена структура викладу. Його основу складає необґрунтоване копіювання чужих думок, фраз і позицій. Студент не вміє робити власні висновки; відсутній творчий компонент даного виду діяльності; практичні вміння і навички мають дуже низький рівень розвитку.

6.4. Форми проведення семестрового контролю та критерії оцінювання

Семестровий контроль із навчальної дисципліни «Лінгвістика комп'ютерного аналізу і обробки інформації» включає змістові блоки «Машинний переклад» та «Комп'ютерна лексикографія» та здійснюється у формі *заліку*. До заліку допускаються студенти, які відвідали всі лекційні, семінарські, практичні заняття, активно працювали, написали всі модульні контрольні роботи та виконали блок самостійної роботи і набрали не меншу за мінімальну кількість балів (35 балів).

6.5. Орієнтовний перелік тем для семестрового контролю (для самоконтролю)

1. Машинний переклад як наукова дисципліна, цілі і завдання. Предмет і об'єкт даної галузі знань.
2. Історія машинного перекладу. Три покоління розвитку.
3. Переваги машинного перекладу і цілі розробки.
4. Види, форми, типи ручного перекладу.

5. Типологія перекладу залежно від ступеня автоматизації: ручний, автоматизований, автоматичний переклад.
6. Стратегії машинного перекладу: послівний переклад, переклад з використанням трансферу, з використанням інтерлінгва, переклад, заснований на глибинному лінгвістичному аналізі, пам'ять перекладів.
7. Загальний алгоритм вирішення задачі машинного перекладу.
8. Синтез тексту на переказною мовою: семантичне уявлення тексту, побудова синтаксичних структур, морфологічний синтез.
9. Структура лінгвістичної бази знань систем машинного перекладу: лінгвістична база даних і лінгвістичний процесор.
10. Компоненти лінгвістичної бази даних систем машинного перекладу.
11. Морфологічний словник в структурі лінгвістичної бази даних.
12. Синтаксичний словник в структурі лінгвістичної бази даних. Семантичний словник в структурі лінгвістичної бази даних.
11. Структура лінгвістичного процесора систем машинного перекладу.
12. Морфологічний і синтаксичний аналіз вихідного тексту в системах машинного перекладу.
13. Семантичний і прагматичний аналіз вихідного тексту в системах машинного перекладу.
14. Лінгвістичні проблеми машинного перекладу на морфологічному рівні. Шляхи вирішення проблем.
15. Лінгвістичні проблеми машинного перекладу на синтаксичному рівні. Шляхи вирішення проблем.
16. Лінгвістичні проблеми машинного перекладу на семантичному рівні. Шляхи вирішення проблем.
17. Лінгвістичні проблеми машинного перекладу на прагматичному рівні. Шляхи вирішення проблем.
18. Синтаксичні трансформації в машинному перекладі.
19. Переклад фразеологізмів у машинному перекладі.
20. Врахування контексту і екстралінгвістичних чинників при вирішенні полісемії.
21. Системи машинного перекладу, засновані на прикладах.
22. Принципи побудови корпусу текстів для систем пам'яті перекладів.
23. Структура систем і алгоритм роботи пам'яті перекладів.
24. Сфера застосування.
25. Контекстологічний словник в структурі систем машинного перекладу.

26. Критерії оцінки якості машинного перекладу.
27. Шкала оцінки якості перекладу.
28. Критерії оцінки якості машинного перекладу. Статистичний критерій.
29. Якість перекладу на різних мовних рівнях.
30. Машинний переклад в контексті досліджень в області штучного інтелекту.
31. Машинний переклад і комп'ютерна лінгвістика.
32. Машинний переклад та інформаційний пошук.
33. Зв'язок машинного перекладу з математичною лінгвістикою і психолінгвістикою.
34. Машинний переклад і теорія перекладу.
35. Машинний переклад в діяльності перекладача.

6.6 Шкала відповідності оцінок

Рейтин -гова оцінка	Оцінка за стобальною шкалою	Значення оцінки
A	90 – 100 балів	Відмінно – відмінний рівень знань (умінь) в межах обов'язкового матеріалу з, можливими, незначними недоліками
B	82-89 балів	Дуже добре – достатньо високий рівень знань (умінь) в межах обов'язкового матеріалу без суттєвих (грубих) помилок
C	75-81 балів	Добре – в цілому добрий рівень знань (умінь) з незначною кількістю помилок
D	69-74 балів	Задовільно – посередній рівень знань (умінь) із значною кількістю недоліків, достатній для подальшого навчання або професійної діяльності
E	60-68 балів	Достатньо – мінімально можливий допустимий рівень знань (умінь)
FX	35-59 балів	Незадовільно з можливістю повторного складання – незадовільний рівень знань, з можливістю повторного перекладання за умови належного самостійного доопрацювання

Рейтин -гова оцінка	Оцінка за стобальною шкалою	Значення оцінки
F	1-34 балів	Незадовільно з обов'язковим повторним вивченням курсу – досить низький рівень знань (умінь), що вимагає повторного вивчення дисципліни

8. Рекомендовані джерела

Основна:

1. Brew C., Dickinson M. Language and Computers. Wiley-Blackwell, 2012. 256 p.
2. Lisbach B., Meyer V. Linguistic Identity Matching. Springer Vieweg, 2013. 261 p.
3. Roland H. Foundations of Computational Linguistics: Man-Machine Communication in Natural Language. Springer-Verlag, 2014. 548 p.
4. Todoric W. Computer Class Vocabulary: Workbook. Independently published, 2020. 25 p.
5. Schittenkopf J. Computer-Assisted Vocabulary Learning: A Constructive Evaluation of Current Off-line and On-line Trainers. 2011. 488 p.
6. Бондар О. І. Сучасна українська мова: Фонетика. Фонологія. Орфоепія. Графіка. Орфографія. Лексикологія. Лексикографія : навч. посіб. Київ : ВЦ “Академія”, 2006. 368 с.
7. Демська О. Вступ до лексикографії. Київ : Вид. дім “Кієво-Могилянська академія”, 2010. 266 с.
8. Кочерган М. П. Вступ до мовознавства : підручник для студентів філологічних спеціальностей вищих навчальних закладів. Київ : Видавничий центр “Академія”, 2001. 368 с.
9. Кровицька О. Українська лексикографія : теорія і практика. НАН України, Ін-т українознавства ім. І. Крип'якевича. Л. : Ін-т українознавства ім. І. Крип'якевича НАН України, 2005. 175 с.
10. Широков В. А. Комп'ютерна лексикографія : Монографія. Київ : Наукова думка, 2011. 351 с.

Додаткова:

1. Бондар О.І. Українське діалектне словникарство початку ХХІ ст.. Вісник Одеського національного університету. Серія філологія. Т. 17. Вип. 4. 2012. С. 5–17.

2. Боярова Л.Г. Термінологічна лексика в загальномовному тлумачному словнику. *Лексикографічний бюлетень*. Київ, 2008. Вип. 17. С. 13–24.
3. Бук С. [Частотний словник](#) офіційно-ділового стилю: принципи укладання та статистичні характеристики лексики. *Лінгвістичні студії*: Зб. наук. праць Донецьк: ДонНУ, 2006. Вип. 14. С. 184–188.
4. Войнов В., Лук'яненко С., Носов К. Можливі комп'ютерні еквіваленти традиційних (паперових) тлумачних словників. *Вісник Нац. ун-ту "Львівська політехніка"*. № 453. 2002. С. 101–105.
5. Дарчук Н.П., Грязнухіна Т.О. [Частотний словник](#) сучасної української публіцистики. *Мовознавство*. 1996. № 5. С. 15–19.
6. Дарчук Н. Параметризована база даних сучасної української мови на основі частотних словників. *Проблеми квантитативної лінгвістики*. Чернівці: Рута, 2005. С. 100–110.
7. Дубічинський В. В. Лексикографія: навч.-метод. посіб. для студ. ден. та заоч. навчання зі спец. "Перекладач" і "Приклад. Лінгвістика". Харків: НТУ "ХПІ", 2012. 68 с.
8. Дубічинський В. В. Українська лексикографія: історія, сучасність та комп'ютерні технології: Навч. Посібник. Національний технічний ун-т Харківський політехнічний ін-т. Харків : НТУ ХПІ, 2004. 164 с.
9. Іваштенко О.О., Сагаровський А.А. Методологія і методика укладання ідіодіалектного словника. *Вісник Харківського національного університету імені В.Н.Каразіна*. Сер.: Філологія. 2009. № 843. Вип. 55. С. 114–119.
10. Русанівський В.М., Широков В.А. Інформаційно-лінгвістичні основи сучасної тлумачної лексикографії. *Мовознавство*. 2002. № 6. С. 7–48.