

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Львівський національний університет імені Івана Франка

Факультет електроніки та комп'ютерних технологій

Катедра оптоелектроніки та інформаційних технологій

Затверджено

на засіданні катедри
оптоелектроніки та інформаційних технологій
факультету електроніки та комп'ютерних
технологій
Львівського національного університету імені
Івана Франка
(протокол № 6 від 24 червня 2025 р.)

Завідувач катедри



Олег КУШНІР

Силабус з навчальної дисципліни

«Комп'ютерна лінгвістика і опрацювання природної мови»,

що викладається в межах ОПП **«Прикладна лінгвістика»**

другого (магістерського) рівня вищої освіти
для здобувачів зі спеціальності **В11 Філологія,**

спеціалізації **В11.10 Прикладна лінгвістика**

Львів 2025 р.

Назва дисципліни	Комп'ютерна лінгвістика і опрацювання природної мови
Адреса викладання дисципліни	Корпус факультету електроніки та комп'ютерних технологій, Львівський національний університет імені Івана Франка: вул. Драгоманова 50, 79005 м. Львів
Факультет та кафедра, за якою закріплена дисципліна	Факультет електроніки та комп'ютерних технологій, кафедра оптоелектроніки та інформаційних технологій
Галузь знань, шифр та назва спеціальності	Галузь знань: В Культура, мистецтво та гуманітарні науки Спеціальність: В11 Філологія Спеціалізація: В11.10 Прикладна лінгвістика
Викладачі дисципліни	Кушнір Олег Степанович, докт. фіз.-мат. наук, професор, професор
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): комп. клас №5, корпус факультету електроніки та комп'ютерних технологій, вул. Драгоманова 50, м. Львів Також можливі онлайн-консультації через Zoom. Для погодження часу онлайн-консультацій слід писати на електронну пошту викладача.
Сторінка дисципліни	https://philology.lnu.edu.ua/course/komp-iuterna-linhvistyka-i-opratsiuvannia-pryrodnoi-movy-2025-2 https://drive.google.com/drive/folders/1Kamy1aZ080cxg0ZwuDX9W8kT2rADpYDF?usp=sharing
Інформація про дисципліну	Дисципліна «Комп'ютерна лінгвістика і опрацювання природної мови» є нормативною дисципліною для освітньої програми «Прикладна лінгвістика», викладається в 1 семестрі в обсязі 4 кредити (за Європейською Кредитно-Трансферною Системою – ECTS).
Коротка анотація дисципліни	Навчальну дисципліну розроблено для одержання здобувачами теоретичних знань з основ комп'ютерної лінгвістики і опрацювання природної мови, формування в них практичних навиків застосування засвоєних знань і методів у розв'язанні прикладних задач, найперше для засвоєння сучасних інформаційних технологій, включно з технологіями штучного інтелекту, і відповідного програмного забезпечення з метою забезпечення високоякісних лінгвістичних досліджень, аналізу текстової інформації, перекладання та текстотворення. Представлено теоретичні основи комп'ютерної лінгвістики, опис відомих програмних продуктів комп'ютерної лінгвістики та обробки природної мови, а також загальні засади використання та створення прикладних програм для автоматизації кількісних філологічних досліджень.
Мета та цілі дисципліни	<i>Метою</i> вивчення дисципліни «Комп'ютерна лінгвістика і опрацювання природної мови» є ознайомлення студентів з теоретичними основами, алгоритмами та методами інформаційних технологій, комп'ютерної лінгвістики, опрацювання природної мови та штучного інтелекту в аналізі текстової інформації, перекладанні та текстотворенні. <i>Цілями</i> дисципліни є формування в студентів практичних навичок, які би дали змогу використовувати засвоєні знання, а також обирати, обґрунтовувати та ефективно застосовувати методи і прикладні програми для опрацювання природної мови.
Література для вивчення дисципліни	Основна література: 1. Волошин В. Г. Комп'ютерна лінгвістика / В. Г. Волошин. – Суми : Університетська книга, 2004. – 382 с.

	<ol style="list-style-type: none"> 2. Кушнір О. С. Основи комп'ютерної лінгвістики (конспект лекцій) / О. С. Кушнір. – Львів: Видавн. Львів. ун-ту, 2023. – 292 с. 3. Математична лінгвістика. Книга 1. Квантитативна лінгвістика / В. В. Пасічник, Ю. М. Щербина, В. А. Висоцька, Т. В. Шестакевич. – Львів: Новий світ – 2000, 2012. – 359 с. 4. Bolshakov I. Computational linguistics. Models, resources, applications / I. Bolshakov, A. Gelbukh. – Mexico : Ciencia de la Computacion, 2004. – 198 p. 5. Delmonte R. Computational Linguistic Text Processing / New York : Nova Science Publishers, 2009. – 382 p. 6. https://www.gutenberg.org/ Додаткова література: 7. Clark A. The Handbook of Computational Linguistics and Natural Language Processing / A. Clark, C. Fox, S. Lappin. – Chichester : John Wiley and Sons, 2010. – 801 p. 8. Jurafsky D. Speech and language processing / D. Jurafsky, J. H. Martin. – New Jersey : Prentice Hall, 2023. – 628 p. 9. Manning C. D. Foundations of statistical natural language processing / Manning C. D., Schutze H. – London : The MIT Press Cambridge, 1999. – 680 p. 10. Repetition characteristic for single texts / O. S. Kushnir, L. B. Ivanitskiy, A. I. Kashuba, M. R. Mostova, V. B. Mykhaylyk // Proc. 5th Int. Conf. on Computational Linguistics and Intelligent Systems (COLINS-2021). Vol. I: Main Conference. Ed. by N. Sharonova et al. – Lviv, Ukraine, April 22-23, 2021. – CEUR Workshop Proc., Vol. 2870. – P. 629–641. 11. Glatch S. Repetition definition: types of repetition in poetry and prose / S. Glatch – “writers.com”, 2024. URL: https://writers.com/repetition-definition 12. Prytula M. Fine-tuning BERT, DistilBERT, XLM-RoBERTa and Ukr-RoBERTa models for sentiment analysis of Ukrainian language reviews // Artificial Intelligence. – 2024. – Vol. 2. – P. 85–97.
Обсяг курсу	Сумарно 120 годин. Із них 16 години лекцій, 16 години практичних занять, а також 88 годин самостійної роботи
Очікувані результати навчання	<p>Після завершення цього курсу студент 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 аналізу текстових даних; - вміти аналізувати моделі комп'ютерної лінгвістики, працювати з відповідними програмними продуктами, застосовувати комп'ютерну техніку, інформаційні технології та програмні продукти для вирішення лінгвістичних задач, користуватися прикладними програмами. <p>Після вивчення даного курсу студенти набудуть такі загальні компетентності:</p> <p>ЗК 1. Здатність спілкуватися державною мовою як усно, так і письмово. ЗК 2. Здатність бути критичним і самокритичним. ЗК 3. Здатність до пошуку, опрацювання та аналізу інформації з різних джерел. ЗК 4. Уміння виявляти, ставити та вирішувати проблеми. ЗК 6. Здатність спілкуватися іноземною мовою. ЗК 7. Здатність до абстрактного мислення, аналізу та синтезу. ЗК 8. Навички використання інформаційних і комунікаційних технологій.</p>

	<p>ЗК 10. Здатність спілкуватися з представниками інших професійних груп різного рівня (з експертами з інших галузей знань/видів економічної діяльності).</p> <p>ЗК 11. Здатність проведення досліджень на належному рівні.</p> <p>ЗК 12. Здатність генерувати нові ідеї (креативність).</p> <p>фахові компетентності:</p> <p>ФК 1. Здатність вільно орієнтуватися в різних лінгвістичних напрямках і школах.</p> <p>ФК 4. Здатність здійснювати науковий аналіз і структурування мовного / мовленнєвого й літературного матеріалу з урахуванням класичних і новітніх методологічних принципів.</p> <p>ФК 5. Усвідомлення методологічного, організаційного та правового підґрунтя, необхідного для досліджень та/або інноваційних розробок у галузі філології, презентації їх результатів професійній спільноті та захисту інтелектуальної власності на результати досліджень та інновацій.</p> <p>програмні результати навчання:</p> <p>ПРН 1. Оцінювати власну навчальну та науково-професійну діяльність, будувати і втілювати ефективну стратегію саморозвитку та професійного самовдосконалення.</p> <p>ПРН 3. Застосовувати сучасні методики і технології, зокрема інформаційні, для успішного й ефективного здійснення професійної діяльності та забезпечення якості дослідження в конкретній філологічній галузі.</p> <p>ПРН 4. Оцінювати й критично аналізувати соціально, особистісно та професійно значущі проблеми і пропонувати шляхи їх вирішення у складних і непередбачуваних умовах, що потребує застосування нових підходів та прогнозування.</p> <p>ПРН 5. Знаходити оптимальні шляхи ефективної взаємодії у професійному колективі та з представниками інших професійних груп різного рівня.</p> <p>ПРН 12. Дотримуватися правил академічної доброчесності.</p> <p>ПРН 13. Доступно й аргументовано пояснювати сутність конкретних філологічних питань, власну точку зору на них та її обґрунтування як фахівцем, так і широкому загалу, зокрема особам, які навчаються.</p> <p>ПРН 16. Використовувати спеціалізовані концептуальні знання з обраної філологічної галузі для розв'язання складних задач і проблем, що потребує оновлення та інтеграції знань, часто в умовах неповної/недостатньої інформації та суперечливих вимог.</p> <p>ПРН 17. Планувати, організовувати, здійснювати і презентувати дослідження та/або інноваційні розробки в конкретній філологічній галузі.</p>
Ключові слова	Інформаційні технології, комп'ютерна лінгвістика, статистична лінгвістика, корпусна лінгвістика, опрацювання природної мови, комп'ютерна лексикографія, машинний переклад, мовні моделі
Формат курсу	Очний
	Проведення лекцій, практичних занять та консультацій для поглибленого розуміння тем
Теми	Див. СХЕМА КУРСУ
Підсумковий контроль, форма	Іспит вкінці семестру
Пререквізити	Для вивчення курсу студенти потребують базових знань з основ мовознавства та інформатики.
Навчальні методи та техніки, які будуть	Інформаційні методи (лекції, бесіди, презентації, ілюстрації, демонстрації); дедуктивні методи на основі узагальнень; евристичні методи

<p>використовуватися під час викладання курсу</p>	<p>(обговорення, проблемні лекції, дискусії); практичні заняття, індивідуальні практичні завдання, інтерактивні методи.</p>
<p>Необхідне обладнання</p>	<p>Комп'ютери з необхідною периферією та доступом до мережі Internet, мультимедіа, корпуси текстів різними природними мовами, вільна та відкрита операційна система Linux, стандартні платформи, навчальні середовища та програмне забезпечення Microsoft Teams, Zoom, Moodle, Grammarly та ін. (у рамках ліцензій Університету), мова програмування, інтерпретатор і стандартні бібліотеки Python, оригінальні авторські програми для попереднього опрацювання текстів і вивчення лінгвістичної статистики, словників текстів і корпусів, довжин слів і речень, ключових слів текстів, фонетичного та силабічного аналізу, повторюваності в текстах, подібності текстів і плагіату тощо (див. схему курсу).</p>
<p>Критерії оцінювання (окремо для кожного виду навчальної діяльності)</p>	<p>Оцінювання проводиться упродовж семестру за 100-бальною шкалою.</p> <p>Бали нараховують за такими видами робіт і з такими співвідношеннями:</p> <ul style="list-style-type: none"> • активність і поточне опитування на лекціях і практичних заняттях: 10% оцінки; максимальна кількість балів 10; • практичні заняття: 30% оцінки; максимальна кількість балів 30; • два теоретичні модулі: 10% оцінки; максимальна кількість балів 10; • іспит: 50% оцінки; максимальна кількість балів 50; <p>Загалом максимум 100 балів.</p> <p>Активність і поточне опитування на лекціях і практичних заняттях оцінюється в 10 балів (максимум 3 бали за заняття):</p> <p>8–10 балів – здобувач проявляє достатню активність на заняттях; висловлює власну думку та цікавиться предметом; правильно й аргументовано відповідає на всі поставлені запитання;</p> <p>5–7 балів – здобувач проявляє на заняттях часткову активність, не достатньо повно відповідає на поставлені запитання, відсутній на деяких заняттях;</p> <p>0–4 бал – здобувач відсутній на багатьох або й більшості занять; відповідає на поставлені додаткові запитання поверхово, фрагментарно.</p> <p>Робота на практичних заняттях оцінюється максимальнo в 30 балів (кожен студент протягом семестру виконує комплексне практичне завдання, яке містить стандартні завдання, а також індивідуальну складову, погоджену з ним на першому занятті):</p> <p>25–30 балів – здобувач в повному обсязі виконав практичне завдання; виконання повне та вичерпне; студент вмiє використовувати на практиці всі методи аналізу за предметом;</p> <p>19–24 бали – здобувач в повному обсязі виконав завдання, але у виконанні допустив незначні помилки; вмiє користуватися всіма підходами та методами, але робить незначні помилки у їхньому застосовуванні;</p> <p>12–18 балів – здобувач виконав завдання частково; допускає помилки, проте вмiє користуватися основними підходами та методами; робить помилки у застосовуванні методів аналізу;</p> <p>6–11 балів – здобувач частково виконав завдання; допускає помилки, не вмiє користуватися більшістю підходів і методів; робить помилки у застосовуванні методів аналізу</p> <p>0–5 балів – здобувач не виконав або майже не виконав поставлених практичних завдань.</p> <p>Кожна контрольна (модульна) робота складається з коротких теоретичних питань, на які треба дати правильні відповіді у стандартному форматі. Правильно виконаною вважається робота, якщо відповіді студента правильна та однозначна. Кожна робота оцінюється за шкалою 5</p>

балів і містить по 5 питань. Правильна відповідь на кожне питання оцінюється в 1 бал.

На іспиті можна здобути максимально **50 балів** (по 10 балів за кожне із 5 екзаменаційних питань, які охоплюють усі матеріали дисципліни). За кожне питання бали нараховують так:

5 балів – здобувач в повному обсязі дав відповідь на питання; відповідь повна, змістовна, логічна; студент знає основні поняття та категорії комп'ютерної лінгвістики та опрацювання природної мови, знає всі особливості практичного застосування відповідних методів аналізу;

4 бали – здобувач дав відповідь, але в ній допустив незначні помилки, знає основні поняття та категорії предмету дисципліни, але робить незначні помилки у застосуванні методів аналізу;

3-2 бали – здобувач дав часткову відповідь; допускає помилки, знає частину понять і категорій за тематикою дисципліни; робить помилки у розумінні та вмінні застосування базових методів;

1 бал – здобувач дав часткову відповідь; у відповіді допускає помилки, не знає більшості основних понять та категорій за тематикою дисципліни, робить помилки у застосуванні методів аналізу

0 балів – здобувач не виконав завдання.

У плані імплементації неформальної освіти здобувач за бажанням може додатково здобути максимально 25 балів за самостійну роботу, пред'явивши сертифікати зі споріднених курсів, наприклад:

«Комп'ютерна лінгвістика»

(<https://www.coursera.org/courses?query=computational%20linguistics>),

«Опрацювання природної мови»

(<https://www.coursera.org/courses?query=nlp>),

«NLTK» (<https://www.classcentral.com/subject/nltk>),

«Опрацювання природної мови на Python»

(<https://www.udemy.com/course/nlp-in-python/?couponCode=ST17MT70725A>),

«Опрацювання природної мови з TensorFlow»

(<https://www.coursera.org/learn/natural-language-processing-tensorflow>),

«OpenNLP», «Stanford CoreNLP», «Lingpipe» тощо.

Контрольні заміри знань проводять у формі стандартних практичних завдань і теоретичних питань.

Академічна доброчесність: Очікується, що практичні та контрольні роботи студентів будуть їхніми оригінальними дослідженнями або міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату або спроб обману.

Відвідування занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та практичні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. Студенти зобов'язані дотримуватися всіх термінів, визначених для виконання видів робіт, передбачених курсом.

Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти також заохочуються до використання іншої літератури та джерел, зокрема наукової літератури, відсутньої серед обов'язкової та рекомендованої.

	<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 студента під час практичних 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, не своєчасного виконання поставлених завдань і т. ін.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p>
<p>Питання до контрольних робіт та іспиту</p>	<p>Перелік питань для проведення підсумкового оцінювання знань студентів:</p> <ol style="list-style-type: none"> 1. Лінгвістика та її структура. 2. Загальне поняття про мову. Мова і мислення. Мова як знакова система. Мова і мовлення. Мовна структура. Рівні мови. 3. Алфавіт. Звуки і букви. Фонологія. Поняття про фонему. 4. Морфологія. Основні поняття. Морфема. Види морфем. 5. Синтаксис. Основні поняття. Словосполучення. Речення. 6. Синтаксичні відносини. Керування. Узгодження. Комунікативна організація речення. 7. Семантика. Семантичні відносини. Парадигматичні та синтагматичні відносини. Семантичні мережі. Фрейми. 8. Лексикографія. 9. Прагматика. Основні поняття. Проблема розуміння текстів. 10. Історична (компаративна) лінгвістика. Синхронія і діяcronія. 11. Основні родини мов. Приклади взаємних впливів. 12. Контрастивна лінгвістика або лінгвістична топологія. 13. Соціолінгвістика та діалектологія. Психолінгвістика. 14. Основні завдання комп'ютерної лінгвістики та опрацювання природної мови. 15. Комп'ютерна лінгвістика. Квантитативна, статистична та математична лінгвістика. 16. Приклади «lingware». 17. Предмет, методи, підходи та проблеми комп'ютерної лексикографії. 18. Загальні положення машинного перекладу. Оцінка та критерії якості перекладу. 19. Типові помилки машинного перекладу. Проблеми та методи вирішення. Сучасний стан комп'ютерного перекладу. 20. Автоматичний аналіз настроїв і думок (sentiment analysis і opinion mining). 21. Комп'ютерне розпізнавання мови та мовлення. Комп'ютерний синтез мовлення. 22. Комп'ютерні стандарти збереження текстових даних. 23. Лінгвістичні знаки. Слова, слововживання, словоформи та лексеми. 24. Препроцесинг текстів. Парсинг. Тегування, стеммінг і лемматизація. 25. Поняття слова. Словоформа та слововживання. Абсолютна та відносна частота слова. Рангові залежності частот слів. Довжина і словник тексту. 26. Комп'ютерне опрацювання окремих текстів. Комплексний лінгвостатистичний аналіз на рівнях букв (звуків), слів, фраз і речень. 27. Поняття про лінгвістичні n-грами. Буквені (символьні) та лексичні n-грами. 28. Довжини складів, слів і речень. 29. Формальний словник тексту. Закон Гіпса. 30. Аналітичні та синтетичні мови. Порівняння словників синтетичних і аналітичних мов. 31. «Словникове багатство» мови. Словникове багатство автора. Параметр TTR. 32. Мови з обмеженим «словником». Ієрогліфічні мови. Лінгвостатистичні особливості східних мов. 33. Електронні корпуси текстів. Типи корпусів. Одночасне статистичне опрацювання корпусів. Закон Гіпса для корпусів. 34. Основи статистики для лінгвістів. Основні терміни, поняття та закони. 35. Методичні основи статистичних досліджень у лінгвістиці. Середнє значення та середнє відхилення випадкової величини. Надійність кількісних даних і висновків. 36. Основи класифікації та порівняння текстів. Лінгвістичні маркери. 37. Встановлення авторства, стилю, жанру за методами комп'ютерної лінгвістики. Основи стилеметрії. Маркери ідіолекту.

	<p>38. Пошук ключових слів текстів. Метод TF-IDF. Ключові слова та інформаційний пошук.</p> <p>39. Інформаційне навантаження текстів. Усереднене семантичне навантаження текстів. Проблеми розрізнення природних і рандомних текстів. Манускрипт Войнич.</p> <p>40. Основи теорії передавання інформації. Формальна інформація, яку містить текст.</p> <p>41. Інформаційна ентропія Шеннона. «Надлишковість» мови. Шуми в передаванні текстової інформації.</p> <p>42. Стислість мови у термінах букв, знаків, слів і речень. Мовленнєва стислість. Порівняння різних мов.</p> <p>43. Перекладацькі трансформації тексту. «Універсалії» перекладознавства.</p> <p>44. Фонетичний розбір і силабіфікація текстів різними мовами за допомогою комп'ютерних програм. Метод сонорності.</p> <p>45. Консонантний коефіцієнт. «Прозорість» і «краса» мови. Порівняння корпусів текстів різними мовами.</p> <p>46. Компаративна комп'ютерна лінгвістика. Українська мова в порівнянні з іншими східнослов'янськими мовами.</p> <p>47. Порівняння характеристик мови, мовлення і словників текстів різними мовами. Комп'ютерні дослідження часової еволюції та діалектів мови.</p> <p>48. Штучні, синтетичні та авторські мови та їхні характеристики. Лінгвістична творчість.</p> <p>49. Штучний інтелект і великі мовні моделі.</p> <p>50. Поняття про моделі природної мови та великі мовні моделі.</p> <p>51. Екскурс у текстотворення. Інформаційні технології в застосуванні до текстотворення.</p> <p>52. Налаштування та тренування мовних моделей.</p>
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

СХЕМА КУРСУ

Тиж.	Тема, план, короткі тези	Форма діяльності (заняття)	Література. Ресурси в Інтернеті	Завдання (практична і самостійна робота), год.	Термін виконання
1, 2	Комп'ютерне опрацювання текстів. Оцифрування текстів. Комп'ютерні стандарти збереження текстових даних. Попередня обробка текстів. Парсинг, стемінг і лемматизація. Лінгвістичні знаки. Абсолютна та відносна частота. Довжина і словник тексту. Комп'ютерне опрацювання окремих текстів. Лінгвостатистичний аналіз на рівнях букв (звуків), слів, фраз і речень. Лінгвістичні n-грами. Довжини складів, слів і речень.	Лекція Практичне заняття	1, 2, 4	Вступне заняття. Академічна доброчесність. 0,5 год. Попереднє опрацювання текстів. 1 год. Закони статистичної лінгвістики для окремих текстів. Статистика лексичних n-грам. 1 год. Визначення середньої довжини слів і речень. 1 год. Самостійна робота 11 год.	1, 2 тиж. семестру
3, 4	Властивості текстів і корпусів. Формальний словник тексту. Закон Гіпса. Аналітичні та синтетичні мови. «Словникове багатство» мови та автора. Параметр TTR. Мови з обмеженим «словником». Ієрогліфічні мови. Лінгвостатистичні особливості східних мов. Електронні корпуси текстів. Типи корпусів. Статистичне опрацювання корпусів.	Лекція Практичне заняття	1, 2, 3, 4	Вивчення словника текстів і параметра TTR. 1 год. Комп'ютерне вивчення корпусів текстів. 1 год. Самостійна робота 11 год.	3, 4 тиж. семестру
5, 6	Статистика для лінгвістів. Основи статистики для лінгвістів. Основні терміни, поняття та закони. Середнє значення та середнє відхилення випадкової величини. Методичні основи статистичних досліджень у лінгвістиці. Надійність кількісних лінгвістичних даних і висновків. Основи класифікації та порівняння текстів. Лінгвістичні маркери. Встановлення авторства, стилю, жанру за методами комп'ютерної лінгвістики. Основи стилеметрії. Маркери ідіолекту.	Лекція Практичне заняття	3, 5, 7, 8	Розрахунки середнього значення та стандартного відхилення лінгвістичних параметрів і статистичний Z-тест. 1,5 год. Самостійна робота 11 год.	5, 6 тиж. семестру
7, 8	Інформаційний пошук. Ключові слова текстів. Методи пошуку ключових слів. Інформаційне навантаження текстів. Усереднене семантичне навантаження текстів. Проблеми розрізнення природних і рандомних текстів. Манускрипт Войнич.	Лекція Практичне заняття	3, 4, 7	Робота з базами лінгвістичних даних. Побудова запитів з урахуванням лінгвістичних міркувань. 1 год. Методи визначення ключових слів у текстах. 1 год. Визначення наявності семантичного навантаження текстів. 1 год. Самостійна робота 11 год.	7, 8 тиж. семестру
9, 10	Теорія інформації. Основи	Лекція Прак-	3, 4, 5, 7	Порівняння мовної та мовленнєвої стислості текстів різними мовами. 1 год.	9, 10 тиж.

	теорії передавання інформації. Формальна інформація, яку містить текст. Інформаційна ентропія. «Надлишковість» мови. Стислість мови у термінах букв, знаків, слів і речень. Мовленнєва стислість. Порівняння різних мов. Перекладацькі трансформації тексту. «Універсалії» перекладознавства.	тичне заняття		Самостійна робота 11 год.	семестру
11, 12	Дослідження природних мов. Фонетичний розбір і силабіфікація текстів за допомогою прикладних програм. Метод сонорності. Консонантний коефіцієнт. «Прозорість» і «краса» мови. Українська мова в порівнянні з іншими східнослов'янськими мовами. Комп'ютерні дослідження часової еволюції та діалектів мови. Штучні, синтетичні та авторські мови та їхні характеристики. Лінгвістична творчість.	Лекція Практичне заняття	2, 3, 4, 6, 7, 8	Фонетичний і силабічний аналіз текстів східнослов'янськими мовами. 1 год. Визначення консонантного коефіцієнта. 1 год. Самостійна робота 11 год.	11, 12 тиж. семестру
13, 14	Подібність текстів. Інформаційні запити на текстові документи. Плагіат. Комп'ютерне встановлення плагіату. Повторюваність у текстах. Коефіцієнт повторюваності на різних лінгвістичних рівнях. Відмінності різних природних і штучних мов. Автоматичний аналіз настрів.	Лекція Практичне заняття	3, 5, 6, 8	Встановлення подібності текстів і плагіату. 1 год. Характеристика повторюваності символів у текстах. 1 год. Самостійна робота 11 год.	13, 14 тиж. семестру
15, 16	Машинний переклад і мовні моделі. Загальні положення машинного перекладу. Оцінка та критерії якості. Типові помилки. Сучасний стан машинного перекладу. Предмет, методи та підходи комп'ютерної лексикографії. Основні проблеми комп'ютерної лексикографії. Поняття про моделі природної мови. Штучний інтелект і великі мовні моделі в текстотворенні. Налаштування та тренування мовних моделей.	Лекція Практичне заняття	2, 3, 4, 6, 8	Програмний пакет машинного перекладу Grammarly. 1 год. Самостійна робота 11 год.	15, 16 тиж. семестру