



## Силабус курсу

### Інтелектуальна обробка тексту та природної мови

Ступінь вищої освіти-магістр

Спеціальність «Комп'ютерні науки»

Освітньо-професійна програма: «Комп'ютерні науки»

Рік навчання: I, Семестр: II

Кредитів: 5 Мова викладання: українська

### Керівник курсу

ППП

к.т.н., доцент. Лип'яніна-Гончаренко Христина Володимирівна

Контактна інформація

xrustya.com@gmail.com, +380968000965

### Опис дисципліни

Дисципліна "Інтелектуальна обробка тексту та природної мови" спрямована на надання знань і навичок для проектування, розробки та застосування алгоритмів і методів підсумкового аналізу природної мови і призначених для автоматизації процесу обробки тексту. Ці підходи використовуються для створення штучних інтелектуальних систем для аналізу та відображення значень з природної мови тексту.

### Структура курсу

Години (лек./пр.)	Тема	Результати навчання	Завдання
2/1	Тема 1. Основи обробки природної мови	Ознайомлення з визначенням природної мови, поняттям лінгвістики, синтаксису і структура мови, а також семантики мови, обробки природної мови, аналітики тексту.	Питання, практична робота
2/2	Тема 2. Python для обробки природної мови	Вміти налаштувати надійне середовище Python. Знати синтаксис і структуру Python. Знати принципи роботи з текстовими даними, основами обробки та аналізу тексту та фреймворками для обробки природної мови.	Питання, практична робота
2/2	Тема 3. Опрацювання та розуміння тексту	Знати основні принципи роботи з токенизації тексту, видалення символів із наголосами, розширення скорочень, видалення спеціальних символів, лематизації, видалення стоп-слів.	Питання, практична робота
2/2	Тема 4. Розробка функцій для подання тексту	Знати принципи роботи з текстовими даними, побудовою текстового корпусу, попередньою обробкою текстового корпусу. Вміти будувати інженерні моделі функцій обробки тексту.	Питання, практична робота
2/1	Тема 5. Класифікація тексту	Знати основні засади класифікація тексту. Вміти проводити класифікацію тексту.	Питання, практична робота
4/1	Тема 6. Конспектування тексту та тематичні моделі	Знати основні засади резюмування тексту та вилучення інформації. Вміти моделювати теми підходом Gensim, Scikit-Learn.	Питання, практична робота
4/2	Тема 7. Подібність тексту та кластеризація	Вміти визначати подібність тексту, аналізувати схожі документи та проводити їх кластеризацію.	Питання, практична робота

4/2	Тема 8. Семантичний аналіз	Вміти проводити семантичний аналіз. Знати підхід WordNet. Вміти створювати тигер NER.	Питання, практична робота
4/1	Тема 9. Аналіз настроїв	Знати основні визначення залежностей в тексті. Вміти проводити попередню обробку та нормалізацію тексту. Вміти проводити класифікацію настрою з контрольованим навчанням та аналіз причинно-наслідкового настрою.	Питання, практична робота
4/1	Тема 10. Перспективи глибокого навчання	Знати основні принципи роботи з вставними реченнями, тенденцій в моделях вставних слів та майбутні сфери застосування моделей на основі глибокого навчання.	Питання, практична робота

### **Літературні джерела**

1. Patel A. A. & Arasanipalai A. U. (2021). Applied natural language processing in the enterprise teaching machines to read write and understand (First). O'Reilly.
2. BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP & Association for Computational Linguistics. (2019). Second blackboxnlp workshop on analyzing and interpreting neural networks for nlp at acl 2019 : florence italy 1 august 2019. Curran.
3. St Mt A. Wiksuana I. Ramantha I. W. & Wiagustini N. L. P. (2019). Working role of technical analysis of slowstochastic indicator and commodity channel index in affecting investment decision in indonesia stock exchange. SSRN. <https://doi.org/10.2139/ssrn.3336466>
4. Agerri R. Aliprandi C. Alkhalifa R. Alzetta C. Angel J. Anselmi G. Appiah Balaji N. N. Aroyehun S. T. Artigas Herold M. F. & Attanasio G. (2021). Evalita evaluation of nlp and speech tools for italian - december 17th 2020. Retrieved February 11 2023 from <http://books.openedition.org/aaccademia/6732>.
5. Heale A. (2021). Generation panic : simple & empowering techniques to combat anxiety. O Books.
6. Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. Fundamentals of artificial intelligence, 603-649.
7. Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. Science China Technological Sciences, 63(10), 1872-1897.
8. Galassi, A., Lippi, M., & Torroni, P. (2020). Attention in natural language processing. IEEE transactions on neural networks and learning systems, 32(10), 4291-4308.
9. Eisenstein, J. (2019). Introduction to natural language processing. MIT press.
10. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., ... & Poon, H. (2021). Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare (HEALTH), 3(1), 1-23.
11. Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. arXiv preprint arXiv:2003.07082..

### **Політика оцінювання**

**Політика щодо дедлайнів та перескладання:** Перескладання модулів відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

**Політика щодо академічної доброчесності:** Списування під час контрольних робіт та екзаменів заборонені (в т.ч. із використанням мобільних пристроїв). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування.

**Політика щодо відвідування:** Відвідування занять є обов'язковим компонентом оцінювання. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбутись в он-лайн формі за погодженням із керівником курсу.

## Оцінювання

Заліковий модуль 1	Заліковий модуль 2	Заліковий модуль 3
30 %	40 %	30 %
1. Виконання та захист практичних робіт (3 роботи по 15 балів) – 45 балів 2. Модульна контрольна робота – 55 балів	1. Виконання та захист практичних робіт (2 роботи по 20 балів) – 40 балів 2. Ректорська контрольна робота – 60 балів	1. Виконання завдань під час тренінгу – 20 балів 2. Написання та захист КПІЗ – 80 балів

### Шкала оцінювання:

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90-100	відмінно	A (відмінно)
85-89	добре	B (дуже добре)
75-84		C (добре)
65-74	задовільно	D (задовільно)
60-64		E (достатньо)
35-59	незадовільно	FX (незадовільно з можливістю повторного складання)
1-34		F (незадовільно з обов'язковим повторним курсом)