



Силабус освітнього компонента

Програма навчальної дисципліни



Аналіз і синтез природньомовної інформації

Шифр та назва спеціальності
113 – Прикладна математика

Інститут
ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма
Інтелектуальний аналіз даних

Кафедра
Комп'ютерна математика і аналіз даних

Рівень освіти
Магістр

Тип дисципліни
Профільна, Вибіркова

Семестр
1

Мова викладання
Українська

Викладачі, розробники



Пікалова Валентина Валеріївна

Valentyna.Pikalova@khp.edu.ua

Доцент кафедри комп'ютерної математики і аналізу даних НТУ «ХПІ»,
к.п.н.

Досвід роботи – 23 років. Автор понад 50 наукових та навчально-методичних праць.

Детальніше про викладача на сайті кафедри

Загальна інформація

Анотація

Розглянуто задачі, теоретичні підходи та методи розв'язання основних типів задач обробки природної мови, на стику штучного інтелекту та математичної лінгвістики. Обробка природної мови (NLP) використовує алгоритми для розуміння людської мови та маніпулювання нею. Ця технологія є однією з найпоширеніших областей машинного навчання. Оскільки штучний інтелект продовжує розширюватися, зростатиме попит на професіоналів, які вміють будувати моделі, які аналізують мовлення та мову, виявляють контекстуальні шаблони та виробляють ідеї з тексту та аудіо. Цей курс надасть вам найсучасніші методи глибокого навчання, необхідні для створення передових систем НЛП. Ви будете готові розробляти програми NLP, які дають відповіді на запитання та аналіз настроїв, створюють інструменти для перекладу мов і узагальнення тексту, і навіть створюють чат-ботів.

Мета та цілі дисципліни

Дисципліна спрямована на набуття необхідних компетентностей в галузі обробки природної мови, машинного навчання та штучного інтелекту.

Формат занять

Лекції, лабораторні роботи, розрахункові завдання, самостійна робота, консультації. Підсумковий контроль – іспит.

Компетентності

- ЗК 4. Здатність виявляти, ставити та вирішувати проблеми у професійній діяльності.
- ЗК 6. Здатність критично оцінювати й переосмислювати накопичений досвід (власний і чужий), аналізувати свою професійну і соціальну діяльність.
- СК 4. Здатність розробляти алгоритми аналізу невизначених великих даних, розробляти відповідні програмні засоби та документацію, проектувати програмні системи, бази даних і знань.
- СК 5. Здатність до проведення математичного і комп'ютерного моделювання та обчислювального експерименту, збору, візуалізації, аналізу та обробки отриманих даних, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів.
- СК 10. Здатність обирати, розробляти, досліджувати та застосовувати математичні моделі і методи для інтелектуального аналізу даних в умовах невизначеності.
- СК 11. Здатність розробляти, досліджувати та застосовувати математичні методи й алгоритми машинного навчання, м'яких обчислень і обчислювального інтелекту для аналізу невизначених даних, прогнозування та прийняття рішень.
- СК 12. Здатність до розробки й експлуатації спеціалізованих програмних засобів інтелектуального аналізу даних, текстів, сигналів і зображень.
- СК 13. Здатність до розробки та експлуатації спеціалізованих програмних засобів обробки великих масивів даних на основі інформаційних технологій розподілених і хмарних обчислень.
- СК 14. Здатність до використання сучасних інформаційних технологій інтелектуального аналізу даних, прогнозування, прийняття рішень, інформаційного пошуку і видобування знань.

Результати навчання

- РН 3. Володіти методами розробки, дослідження та застосування математичних моделей складних об'єктів і процесів, у тому числі із застосуванням методів обчислювального інтелекту.
- РН 4. Уміти поєднувати методи математичного і комп'ютерного моделювання з неформальними процедурами експертного аналізу для пошуку оптимальних рішень.
- РН 7. Уміти застосовувати сучасні технології програмування та розроблення програмного забезпечення, програмної реалізації чисельних і символічних алгоритмів.
- РН 8. Уміти застосовувати у практичній роботі спеціалізовані програмні продукти і програмні системи комп'ютерної математики, аналізу великих даних тощо.
- РН 11. Демонструвати навички професійного спілкування, усної та письмової комунікації українською мовою і принаймні ще однією з європейських мов.
- РН 13. Знати і розуміти методи розв'язання математичних задач інтелектуального інформаційного пошуку та видобування знань.
- РН 14. Уміти застосовувати наявні існуючі і розробляти нові алгоритми та програмні засоби для статистичного й інтелектуального аналізу невизначених даних.
- РН 15. Уміти застосовувати наявні існуючі і розробляти нові алгоритми та програмні засоби обробки даних, текстів, сигналів і зображень.
- РН 16. Уміти застосовувати сучасні інформаційні технології та програмне забезпечення для обробки великих масивів даних на основі розподілених і хмарних сервісів.

Обсяг дисципліни

Загальний обсяг дисципліни 150 год. (5 кредитів ECTS): лекції – 32 год., лабораторні роботи – 32 год., самостійна робота – 86 год.

Передумови вивчення дисципліни (пререквізити)

Бакалаврський рівень підготовки.

Особливості дисципліни, методи та технології навчання

При викладанні даної дисципліни використовуються такі методи навчання і викладання, як гейміфікація та peer-to-peer. В процесі навчання використовуються системи LMS (learning management systems).

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Класифікація та векторні простори в НЛП
Тема 2. Векторні просторові моделі
Тема 3. Імовірнісні моделі в NLP
Тема 4. Позначення частин мови (POS).
Тема 5. Моделі мови N-грам
Тема 6. Моделі послідовності в NLP
Тема 7. Моделі генерації мови
Тема 8. Сіамські мережі
Тема 9. Моделі уваги в НЛП
Тема 10. Нейронний машинний переклад з увагою (Attention) |

Теми практичних занять

| Практичні заняття в рамках дисципліни не передбачені. |

Теми лабораторних робіт

Тема 1. Логістична регресія для аналізу настроїв твітів
Тема 2. Наївний Байєс для аналізу настроїв твітів
Тема 3. Вбудовування слів і хешування з урахуванням локалізації для машинного перекладу
Тема 4. Автоматичне виправлення за допомогою мінімальної відстані редагування (Minimum Edit Distance)
Тема 5. Word2Vec і стохастичний градієнтний спуск
Тема 6. Почуття за допомогою нейронних мереж
Тема 7. Розпізнавання іменованих сутностей (NER)
Тема 8. Підведення підсумків за допомогою моделей трансформерів (Transformer Models)
Тема 9. Запитання-відповіді з моделями трансформерів
Тема 10. Розробка чат-бота з моделлю реформатора |

Самостійна робота

| Курс передбачає виконання індивідуальних завдань, результат розв'язання яких перевіряється автоматично засобами LMS та контролюються та оцінюються викладачами. Студентам також рекомендуються додаткові матеріали (відео, статті) для самостійного вивчення. |

Література та навчальні матеріали

Основна література

1. Кононова К. Ю. Машинне навчання. – Харків: ХНУ ім. В. Н. Каразіна, 2020. – 301 с.
https://moodle.znu.edu.ua/pluginfile.php/593075/mod_folder/intro/Базовий%20підручник%202%20%208Кононова%20К.%20Ю.%20Машинне%20навчання%20-%20методи%20та%20моделі%29.pdf
2. G. James, D. Witten, T. Hastie, R. Tibshirani. An Introduction to Statistical Learning with Applications in R. – New York: Springer, 2015. – 440 с. – ISBN 978-146-147-138-7.
3. P. Harrington. Machine Learning in Action. – Shelter Ireland, NY: Manning, 2012. – 400 с. – ISBN 978-966-423-364-1.
4. О. В. Бісикало, Формальні методи образного аналізу та синтезу природно- мовних конструкцій. Монографія, Вінниця: ВНТУ, 2013, – 316 с. – ISBN 978-966-641-528-1.
5. Principles of Natural Language Processing by Susan McRoy is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted. – ISBN 978-1-7376595-0-1, EPub – ISBN 978-1-7376595-1-8.
<https://uwm.pressbooks.pub/naturallanguage/open/download?type=pdf>

Додаткова література

6. A. Burkov. Machine Learning Engineering. – True Positive Inc., 2020. – 310 p. – ISBN 978-199-957-957-9.

7. М. Kubat. An Introduction to Machine Learning. - Springer Cham, 2021. – 458 p. – ISBN 978-303-081-935-4.

8. S.T. Wierzchoiń, M.A. Kłopotek. Modern Algorithms of Cluster Analysis - Springer, Cham, 2018. - 421 p. – ISBN 978-3-319-69308-8.

Інтернет-ресурси

9. https://en.wikipedia.org/wiki/Natural_language_processing

10. <https://www.deeplearning.ai/courses/natural-language-processing-specialization/>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

100% підсумкової оцінки складаються з результатів оцінювання у вигляді іспиту (40%) та поточного оцінювання (60%).

Іспит: письмове завдання (2 теоретичних і задача) та усна доповідь.

Поточне оцінювання: оцінки за лабораторні роботи, 2 контрольні роботи та розрахункове завдання

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>.

Погодження

Силабус погоджено

Дата погодження, підпис
31.08.2023 р.



Завідувач кафедри
Олена АХІЄЗЕР

Дата погодження, підпис
31.08.2023 р.



Гарант ОП
Олексій ГАЛУЗА