Затверджено Вченою Радою ОНУ
імені І.І. Мечникова
від “20”грудня 2016 р. № 4

Одеський національний університет імені І.І. Мечникова
Факультет математики, фізики та інформаційних технологій
Кафедра математичного забезпечення комп’ютерних систем

“ЗАТВЕРДЖУЮ”
Проректор з науково-педагогічної роботи
_____
“_____” _____ 20___ р.

НАВЧАЛЬНА ПРОГРАМА ДИСЦИПЛІНИ

ВБ11 Дисципліна спеціалізації №1 «Застосування Python-засобів для обробки текстів природної мови»
(назва навчальної дисципліни)

Рівень вищої освіти _____бакалавр_____

Спеціальність _____126 – Інформаційні системи та технології_____
(код і назва спеціальності (тей)

2019

Розробник:
Пенко В.Г., к.т.н., доцент кафедри МЗКС

Навчальна програма затверджена на засіданні кафедри математичного забезпечення комп'ютерних систем

Протокол № 1 від. "30" 08 2019 р.

Завідувач кафедри _____ (підпис) (Є.В. Малахов)
(прізвище та ініціали)

Обговорено та рекомендовано до затвердження навчально-методичною комісією (НМК) з ІТ факультету МФІТ

Протокол № 1 від. "23" 09 2019 р.

Голова НМК _____ (підпис) (О.В. Савастру)
(прізвище та ініціали)

Вступ

Навчальна програма дисципліни спеціалізації №1 «Застосування Python-засобів для обробки текстів природної мови» складена відповідно до освітньо-професійної програми підготовки бакалаврів спеціальності 126 – Інформаційні системи та технології (версія 2019 р.).

Предметом засоби автоматизації обробки тексів природною мовою та вирішення базових завдань, повязаних з такими текстами.

Місце дисципліни і значення в структурі навчального процесу.

Вивчення курсу та виконання лабораторних робіт базується на знанні програмування, в першу чергу об'єктно-орієнтованому, що викладаються у дисциплінах «Алгоритмізація та програмування», «Об'єктно-орієнтоване програмування».

Програма навчальної дисципліни складається з таких змістових модулів:

1. Обробка текстів в Python.
2. Отримання доступу до корпусів текстів і лексичних ресурсів.

1. Мета та завдання навчальної дисципліни

Метою курсу є вивчення практичних засобів застосування мови та бібліотек Python для вирішення базових завдання обробки текстів на природній мові.

Завдання:

– освоєння навичок застосування мови Python для розробки програм загального призначення;

– знайомство з проблематикою обробки текстів на природній мові;

– застосування можливостей спеціалізованих пакетів для підвищення ефективності обробки природної мови.

Процес вивчення дисципліни спрямований на формування елементів наступних **компетентностей**:

1) загальних:

КЗ 2. Здатність застосовувати знання у практичних ситуаціях.

КЗ 3. Здатність до розуміння предметної області та професійної діяльності.

КЗ 5. Здатність вчитися і оволодівати сучасними знаннями.

КЗ 6. Здатність до пошуку, оброблення та узагальнення інформації з різних джерел.

2) спеціальних (фахових):

КС 1. Здатність аналізувати об'єкт проектування або функціонування та його предметну область.

КС 6. Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання функціональних завдань та обов'язків.

КС 13. Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень .

Програмні результати навчання:

ПР 2. Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв'язанні задач проектування і використання інформаційних систем та технологій.

ПР 3. Використовувати базові знання інформатики й сучасних інформаційних систем та технологій, навички програмування, технології безпечної роботи в комп'ютерних мережах, методи створення баз даних та інтернет-ресурсів, технології розроблення алгоритмів і комп'ютерних програм мовами високого рівня із застосуванням об'єктно-орієнтованого програмування для розв'язання задач проектування і використання інформаційних систем та технологій.

ПР 6. Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності.

Очікувані результати навчання. У результаті вивчення навчальної дисципліни студент повинен

***знати**: основні можливості мови Python для розробки программного забезпечення обробки текстів; основні можливості спеціалізованих пакетів для обробки*

текстів на природній мові; основні типи завдань, що стосуються обробки текстів на природній мові; особливості корпусо-орієнтованого підходу до обробки текстів природної мови.
***вміти**: розробляти програмне забезпечення, що виконує базові операції з текстами; використовувати спеціалізовані пакети Python для підвищення єфективності вирішення базових завдан обробки текстів на природній мові; застосовувати декілька різновидів мовних корпусів в якості ресурсу для вирішення завдань обробки текстів на природній мові.*

2. Зміст навчальної дисципліни

Змістовний модуль 1 Обробка текстів в Python.
Тема 1. Обчислення над мовою - проста статистика.
Література: [2, 6].
Тема 2. Автоматизація розуміння природної мови.
Література: [3, 6].
Тема 3. Отримання доступу до корпусів текстів і лексичних ресурсів.
Література: [2, 5].
Тема 4. Використання лексичних ресурсів.
Література: [1, 6, 8].

Змістовний модуль 2 Отримання доступу до корпусів текстів і лексичним ресурсів.
Тема 1. Корпус WordNet.
Література: [6, 9].
Тема 2. Доступ до тексту з Web і до локального тексту.
Література: [1, 6].
Тема 3. Трубопровід NLP. Реалізація окремих етапів.
Література: [1, 3, 4, 6].
Форма контролю: **КО** –контрольне опитування (поточне)

3. Рекомендована література

3.1. Основна література

1. Hobson Lane, Howard Cole, Hannes Max Hapke Natural Language Processing in Action: Understanding, analyzing, and generating text with Python: Manning Shelter Island, 2019. - 576p.
2. А. Васильев Програмування мовою Python Навчальна книга – Богдан, 2019. - 514 с.
3. N. Indurkhya, F.J.Damerau Handbook of Natural Language Processing: A Chapman & Hall Book/CRC Machine Learning & Pattern Recognition Series , 2010. - 676 p.

3.2. Допоміжна література

4. C. D. Manning, H. Schutze Foundations of Statistical Natural Language Processing The MIT Press Cambridge, Massachusetts London, England, 1999. – 717 p.

5. Хайрова Н. Ф., Мамырбаев О. Ж., Петрасова С. В., Мухсина К. Ж. Современные технологии обработки текстовых данных на базе пакета NLTK Python : учеб. пособ. Харьков : НТУ «ХПИ», 2020. 134 с.
6. D. Jurafsky, J. H. Martin Speech and Language Processing - Prentice Hall, Englewood Cliffs, New Jersey 2008. – 975 p.

3.3. Електронні інформаційні ресурси

7. Natural Language Toolkit – Режим доступу: https://www.nltk.org/
8. Natural Language Processing with Python – Режим доступу: https://www.nltk.org/book/
9. Project Gutenberg – Режим доступу: https://www.gutenberg.org/
10. WordNet:A Lexical Database for English – Режим доступу: https://wordnet.princeton.edu/

11. Форма підсумкового контролю успішності навчання

Другий семестр – залік.

12. Методи діагностики успішності навчання

Поточне опитування, оцінка виконання лабораторних та самостійних завдань, підсумкове опитування.

Примітки:

1. Програма навчальної дисципліни розробляється вищим навчальним закладом. Програма навчальної дисципліни визначає місце і значення навчальної дисципліни, її загальний зміст та вимоги до знань і вмінь.
2. Програма навчальної дисципліни розробляється на основі освітньо-професійної програми.
3. Форма призначена для складання робочої програми навчальної дисципліни.
4. Вищими навчальними закладами можуть вноситися зміни до форми та змістового наповнення «Програми навчальної дисципліни» залежно від специфіки та профілю вищого навчального закладу.
5. Формат бланка – А4 (210×297 мм).