


**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХЕРСОНСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК, ФІЗИКИ ТА МАТЕМАТИКИ
КАФЕДРА ІНФОРМАТИКИ, ПРОГРАМНОЇ ІНЖЕНЕРІЇ ТА
ЕКОНОМІЧНОЇ КІБЕРНЕТИКИ**

ЗАТВЕРДЖЕНО

на засіданні кафедри інформатики, програмної
інженерії та економічної кібернетики

протокол від *07.09.2020* р. № *1*

завідувач кафедри

 (проф. В. Песчаненко)

**СИЛАБУС ОСВІТНЬОЇ КОМПОНЕНТИ
ТЕОРІЯ РОЗДІЛЬНОЇ КОНЦЕПТУАЛЬНОЇ МОДЕЛІ**

Освітня програма: Філологія (прикладна лінгвістика)

Спеціальність: 035 Філологія

другого (магістерського рівня)

Галузь знань: 03 Гуманітарні науки

Херсон 2020

1. Опис курсу

Назва освітньої компоненти	Теорія роздільної концептуальної моделі
Тип курсу	Обов'язкова компонента
Ступінь вищої освіти	Другий (магістерський) рівень освіти
Кількість кредитів/годин	4 кредити / 120 год.
Семестр	II семестр
Викладач	Полторацький М. Ю. https://orcid.org/0000-0001-9861-4438
Посилання на сайт	https://sites.google.com/site/maksympoltorackiy
Контактний телефон	0506686238
Email викладача:	mpoltorackiy@gmail.com
Графік консультацій	П'ятниця, 16:00-17:00, ауд. 503 або за призначеним часом
Методи викладання	лекційні заняття, лабораторні роботи, кейси, презентації, тестові завдання, індивідуальні завдання
Форма контролю	Екзамен

2. Анотація дисципліни: навчальна дисципліна розрахована на години лекційних і практичних занять з метою обговорення проектних робіт по автоматизації та обробки текстового масиву інформації засобами мов програмування. Вивчається комплексне використання методології, інструментальних засобів обробки текстового масиву інформації; виробка

навиків програмування; методики обробки текстового масиву інформації; а також мови програмування для автоматизації обробки текстової інформації; методи обробки текстової інформації; інструментарій для автоматизації обробки текстової інформації; обробка та візуалізація текстового масиву інформації.

3. Мета та завдання дисципліни:

Мета курсу:

Дисципліна «Теорія роздільної концептуальної моделі» повинна забезпечити формування загальнокультурних і професійних компетенцій в частині виконання проектних робіт по автоматизації та обробки текстового масиву інформації засобами мов програмування.

Завдання курсу:

- комплексне використання методології, інструментальних засобів обробки текстового масиву інформації;
- виробка навиків програмування
- вивчення методики обробки текстового масиву інформації.

У результаті вивчення навчальної дисципліни студент повинен

знати:

- використання мов програмування для автоматизації обробки текстової інформації;
- методи обробки текстової інформації

вміти:

- застосовувати інструментарій для автоматизації обробки текстової інформації;
- здійснювати обробку та візуалізацію текстового масиву інформації.

4. Програмні компетентності та результати навчання

Після успішного завершення дисципліни здобувач формуватиме наступні програмні компетентності та результати навчання:

ІК. Здатність розв'язувати складні задачі і проблеми в галузі лінгвістики та інформаційних технологій в процесі професійної діяльності або навчання, що передбачає проведення досліджень та/або здійснення інновацій та характеризується невизначеністю умов і вимог.

Загальні компетентності:

ЗК-1. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК-3. Здатність до пошуку, опрацювання та аналізу інформації з різних джерел.

ЗК-7. Здатність до абстрактного мислення, аналізу та синтезу.

Фахові компетентності:

ФК-1. Здатність вільно орієнтуватися в різних лінгвістичних напрямках і школах.

ФК-6. Здатність професійно застосовувати поглиблені знання з обраної філологічної спеціалізації для вирішення професійних завдань.

ФК-7. Здатність вільно користуватися спеціальною термінологією в обраній галузі філологічних досліджень.

ФК-10. Здатність застосовувати набуті знання у галузі інформаційних технологій у професійній, науково-дослідній та перекладацькій діяльності, зокрема у процесі створення електронних корпусів текстів, баз даних, електронних засобів навчального призначення.

ФК-11. Здатність використовувати на практиці основи лінгвістичного моделювання, системи автоматизованого перекладу, мов програмування.

Програмні результати навчання:

ПРН-2. Упевнено володіти державною та іноземною мовами для реалізації письмової та усної комунікації, зокрема в ситуаціях професійного й наукового спілкування.

ПРН-3. Застосовувати сучасні методики і технології, зокрема інформаційні, для

успішного й ефективного здійснення професійної діяльності та забезпечення якості дослідження в конкретній філологічній галузі.

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.

ПРН-18. Впроваджувати елементи теоретичного та експериментального дослідження, виконувати базові дослідницькі лінгвістичні/перекладознавчі завдання з урахуванням елементів програмування й автоматичної обробки тексту.

ПРН-19. Описувати основні засадничі принципи прикладної та експериментальної лінгвістики, класифікувати формальні методи прикладної лінгвістики, причини, перебіг і наслідки процесів, що відбуваються з огляду на математичну логіку; розуміти значущість комп'ютерних технологій в мовознавстві.

2. Структура курсу

Кількість кредитів/годин	Лекції (год.)	Практичні заняття (год.)	Самостійна робота (год.)
4 кредити / 120 годин	20 год.	20 год.	80 год.

3. Технічне й програмне забезпечення/обладнання

Комп'ютерні аудиторії кафедри.

Програмне забезпечення:

- Python
- Git

4. Політика курсу

Для успішного складання підсумкового контролю з дисципліни вимагається 100% відвідування очне або дистанційне відвідування всіх лекційних занять. Пропуск понад 25% занять без поважної причини буде оцінений як FX.

Високо цінується академічна доброчесність. До всіх студентів освітньої програми відбувається абсолютно рівне ставлення. Навіть окремий випадок порушення академічної доброчесності є серйозним проступком, який може призвести до несправедливого перерозподілу оцінок і, як наслідок, загального рейтингу студентів. Мінімальне покарання для студентів, яких спіймали на обмані чи плагіаті під час тесту чи підсумкового контролю, буде нульовим для цього завдання з послідовним зниженням підсумкової оцінки дисципліни принаймні на одну літеру. Будь ласка, поставтесь до цього питання серйозно та відповідально.

5. Схема курсу

Модуль 1. Аналіз текстової інформації засобами Pullenti, ML та IMS

Тема 1. Огляд сучасних систем для морфологічного та семантичного аналізу тексту (Тижд. 1, 2; лк - 4 год, практ. зан. - 4 год., сам. роб. - 16 год.)

1. NLTK, ОРФО, RCO Morphology SDK
2. Модулі машинної морфології AskNet Morphology, Pullenti SDK, ЕТАП- 3, Stanford Parser, Семантичне дзеркало, LingPipe, МетаФраз.

Тема 2. Python.Pullenti, як інструмент для аналізу текстової інформації (Тижд. 3, 4; 4 лк - 4 год, практ. зан. - 4 год., сам. роб. - 16 год.)

1. Вступ до SDK Python.Pullenti.
2. Модель текстового токена для Python.Pullenti.

3. Модель іменованої сутності для Python.Pullenti.

Тема 3. Аналіз тексту засобами Pullenti (Тижд. 5, 6, 7; лк - 4 год, практ. зан. - 4 год., сам. роб. - 16 год.)

1. Використання процесорів в Python.Pullenti.
2. Аналізатори в Python.Pullenti.(Ner.Date.DateAnalyzer, Ner.Money.MoneyAnalyzer, Ner.Keyword.KeywordAnalyzer, Ner.GeoAnalyzer)

Модуль 2. Аналіз текстової інформації засобами Pullenti, ML та IMS

Тема 4. Аналіз тексту засобами Pullenti (Тижд. 8, 9, 10; лк - 4 год, практ. зан. - 4 год., сам. роб. - 16 год.)

1. Аналізатори в Python.Pullenti. (Ner.Address.AddressAnalyzer, Ner.OrganizationAnalyzer, Ner.PersonAnalyzer)
2. Приклади використання аналізаторів Python.Pullenti для обробки текстової інформації

Тема 5. Використання методів та засобів ML та IMS для аналізу текстового масиву інформації (Тижд. 11, 12, 13; лк - 4 год, практ. зан. - 4 год., сам. роб. - 16 год.)

1. Використання методів обробки юридичної текстової інформації засобами Python.Pullenti для автоматичного створення IMS проекту.

9. Система оцінювання та вимоги: форма (метод) контрольного заходу та вимоги до оцінювання програмних результатів навчання

Модуль 1. Аналіз текстової інформації засобами Pullenti, ML та IMS
кількість балів за цей модуль

Модуль 2. Аналіз текстової інформації засобами Pullenti, ML та IMS а кількість балів за цей модуль

Форма (метод) контрольного заходу, критерії оцінювання та бали

Лабораторні роботи – 25 балів (по 5 балів за 5 лабораторні роботи)

Практичне завдання – 25 балів.

Підсумковий тест за дисципліну – 10 балів.

10. Список рекомендованих джерел (наскрізна нумерація)

1. Виолентов И. А. Анализ подходов и инструментов для извлечения именованных сущностей // *Труды 20-ой Международной научно-технической конференции «Информационные системы и технологии»*. 2014. Р. 227.
2. Огарок А. Л. Метод полного лингвистического анализа неструктурированной текстовой информации // *Информатизация и связь*. 2018. №1. Р. 91-99.
3. Солошенко А. Н., Орлова Ю. А., Розалиев В. Л. Автоматизированный анализ новостного потока из сети Интернет с учетом семантики текстов // *Известия Волгоградского государственного технического университета*. 2014. № 25. Р. 80-85.
4. Baldwin B., Dayanidhi K. Natural language processing with Java and LingPipe Cookbook. Packt Publishing Ltd, 2014.
5. Khoroshevsky V. F. Ontology Driven Multilingual Information Extraction and Intelligent Analytics. 2010.
6. Loper E., Bird S. NLTK: the natural language toolkit. arXiv preprint cs/0205028, 2002.

7. Tefera A., Assabie Y. Automatic Construction of Amharic Semantic Networks From Unstructured Text Using Amharic WordNet // *Proceedings of the Seventh Global Wordnet Conference*. 2014. P. 172-177.