

Черкаський державний технологічний університет
Факультет інформаційних технологій і систем

«ЗАТВЕРДЖУЮ»

Голова вченої ради
факультету

Протокол № 5
«17» лютого 2020



СИЛАБУС

навчальної дисципліни

«Технології обробки природо мовної інформації»

Шифр за ОПП – ВІШІБ8

Освітній рівень - бакалаврський

Галузь знань - 12 – інформаційні технології

Спеціальність - 126 – інформаційні системи та технології

Освітня програма - «Web-технології, Web-дизайн»

2019 - 2020 навчальний рік

Силабус навчальної дисципліни «Технології обробки природо мовної інформації»

(назва навчальної дисципліни)

підготовки здобувачів освітнього ступеня «бакалавр» за спеціальністю 126 – Інформаційні системи та технології, освітня програма «Web-технології , Web-дизайн» - 13 стор.

Силабус складений на основі програми навчальної дисципліни «Технології обробки природо мовної інформації», шифр (за ОПП) – ВППБ8.

Розробник силабусу:

Тарасенко Ярослав Володимирович, к.т.н., асистент кафедри ІТП

(ПІБ, наук.ст., вчене зв., посада НПП кафедри, що розробив силабус)


Силабус затверджений на засіданні кафедри інформаційних технологій проектування

Протокол № 8 від «10» січня 2020 року

Обговорено та рекомендовано до затвердження методичною комісією факультету інформаційних технологій і систем

«14» лютого 2020 р., протокол № 4

Голова методичної комісії
факультету інформаційних технологій і систем



підпис

/А.Р. Карапетян/

ПІБ

1. ІНФОРМАЦІЯ ПРО ВИКЛАДАЧА

Прізвище, ім'я, по батькові	Тарасенко Ярослав Володимирович
Науковий ступінь	к.т.н.
Наукове звання	-
Посада	асистент
Місце роботи	Черкаський державний технологічний університет
Адреса кафедри	18006, м. Черкаси, бул. Шевченка 460, каб. 603-1 корпус
Контактний телефон	(0472)51-15-86
Профайл викладача	https://chdtu.edu.ua/fitis/kitp/staff/item/13114-tarasenko-yaroslav-volodymyrovych
e-mail:	ya.tarasenko@chdtu.edu.ua
Профайл дисципліни	http://fitis.moodle.chdtu.edu.ua/course/view.php?id=574
Розклад консультацій	

2. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Галузь знань, спеціальність, освітня програма, освітній рівень	Загальні характеристики		Навчальне навантаження з дисципліни	
			денна форма навчання	заочна форма навчання
<i>Галузь знань</i> 12 – інформаційні технології	Вибіркова		Курс підготовки:	
			4-й	
<i>Спеціальність</i> 126 – інформаційні системи та технології	Загальна кількість кредитів ЄКТС	4	Семестр підготовки:	
	Загальна кількість годин	120	7-й	
<i>Освітня програма</i> «Web-технології, Web-дизайн»	Кількість аудиторних годин	48	Лекції	
	Кількість годин самостійної роботи	72	16 год.	
			Практичні, семінарські	
<i>Освітній рівень</i> бакалаврський	Мова навчання - українська		Лабораторні	
			32 год.	
			Самостійна робота	
			72 год	
			Форма підсумкового контролю	
			Залік	

3. МЕТА І ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета викладання дисципліни	Навчання майбутніх фахівців особливостей використання програмних засобів та створення власних програм з метою обробки та маніпуляції природо мовною інформацією шляхом застосування формальної граматики та семантики.
Завдання вивчення дисципліни	Розвиток у здобувачів конструктивної грамотності, що забезпечує оволодіння навичками застосування та розробки інформаційних технологій для прикладного застосування статистичних моделей природніх мов з метою використання в комп'ютерних системах обробки природньої мови та у системах природо мовного діалогу.

4. РЕЗУЛЬТАТИ НАВЧАННЯ

№ з/п	Результати навчання
1	Здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші).
2	Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.
3	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.
4	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.

5. ПРЕРЕКВІЗИТИ

Алгоритмізація та програмування, об'єктно-орієнтоване програмування, бази даних та знань, теорія ймовірності математична статистика, методи та системи штучного інтелекту, інтелектуальний аналіз даних, людино-машинна взаємодія.

6. ПОСТРЕКВІЗИТИ

Прикладні інтелектуальні системи обробки даних, технології та системи підтримки прийняття рішень.

7. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

<p style="text-align: center;">Змістовий модуль №1</p> <p style="text-align: center;"><i>Основні моделі методи та засоби комп'ютерної обробки природо мовної інформації.</i></p>
<p>Тема 1 <i>Базові принципи комп'ютерної обробки природної мови.</i></p>
<p><i>1.1. Мовна компетенція комп'ютерів. 1.2. Область застосування технологій обробки природо мовної інформації. 1.3. Інструментальні програмні засоби обробки тексту природної мови. 1.4. Узагальнення властивостей обробки різних мов.</i></p>
<p>Тема 2 <i>Математичні методи аналізу природної мови.</i></p>
<p><i>2.1. Статистичні моделі та дискретний аналіз природних мов. 2.2. Застосування неklasичних логік в обробці природної мови. 2.3. Методи автоматизованого морфологічного та синтаксичного аналізу. 2.4. Методи дослідження семантики тексту природної мови. 2.5. Реалізація математичних методів засобами PHP та Python.</i></p>
<p>Тема 3 <i>Формальна граматики природної мови.</i></p>
<p><i>3.1. Особливості формальної мови. 3.2. Поняття контекстно-вільної та контекстно-залежної граматики. 3.3. Різновиди генеративної граматики. Трансформаційна граматики. 3.4. Основні аспекти граматики зв'язків.</i></p>
<p>Тема 4 <i>Формальна семантика природної мови.</i></p>
<p><i>4.1. Онтологічний підхід в дослідженні семантики тексту природної мови. 4.2. Методи формальної семантики. 4.3. Інтенціональна логіка та семантика Монтегю. 4.4. Дискурсний аналіз та формальна семантика.</i></p>
<p style="text-align: center;">Змістовий модуль №2</p> <p style="text-align: center;"><i>Прикладне застосування методики автоматизованої обробки природомовних об'єктів.</i></p>
<p>Тема 1 <i>Алгоритми пошуку в тексті.</i></p>
<p><i>1.1. Прямий пошук. 1.2. Алгоритм Кнута, Моріса і Прата. 1.3. Алгоритм Бойєра-Мура. 1.4. Семантичний пошук.</i></p>
<p>Тема 2 <i>Проектування, розробка та використання засобів реалізації машинного перекладу.</i></p>
<p><i>2.1. Сучасний рівень технологій автоматизованого перекладу. 2.2. Статистичний та гібридний методи в проектуванні та розробці спеціалізованих інформаційних системах автоматизованого перекладу. 2.3. Підходи до машинного перекладу. 2.4. Класифікація систем машинного перекладу та вибір їх представників.</i></p>
<p>Тема 3 <i>Автоматизоване реферування текстової інформації.</i></p>

3.1. Методи та алгоритми скорочення текстової інформації. 3.2. Автоматизовані системи реферування. 3.3. Етапи реферування текстової інформації. 3.4. Визначення межі семантичного стиснення.

Тема 4 Системи природо мовного діалогу.

4.1. Корпус текстів. 4.2. Мовні діалогові системи та моделі діалогу. 4.3. Особливості розробки діалогової системи з використанням корпусу текстів. 4.4. Лінгво-комбінаторне моделювання погано формалізованих систем.

8. ТЕМАТИЧНИЙ ПЛАН НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

№ теми	Назва модулів і тем	Форми організації навчання, кількість годин						Література, інформаційні ресурси
		Денна форма			Заочна форма			
		Лекції	Практичні, лабораторні роботи	Самостійна робота	Лекції	Практичні, лабораторні роботи	Самостійна робота	
Змістовий модуль №1. Основні моделі методи та засоби комп'ютерної обробки природо мовної інформації.								
1	Базові принципи комп'ютерної обробки природної мови.	2	4	6	-	-	-	1, 2, 3, 5
2	Математичні методи аналізу природної мови.	2	4	8	-	-	-	1, 2, 5, 6, 7
3	Формальна граматики природної мови.	2	4	10	-	-	-	1, 2, 4, 5
4	Формальна семантика природної мови.	2	4	14	-	-	-	2, 4, 6, 7
Змістовий модуль №2. Прикладне застосування методик автоматизованої обробки природомовних об'єктів.								
5	Алгоритми пошуку в тексті.	2	4	8	-	-	-	1, 3, 5, 6
6	Проектування, розробка та використання засобів реалізації машинного перекладу.	2	4	8	-	-	-	1, 3, 5, 8
7	Автоматизоване реферування текстової інформації.	2	4	10	-	-	-	1, 2, 3, 7
8	Системи природо мовного діалогу.	2	4	8	-	-	-	1, 3, 6, 8
	Разом	16	32	72	-	-	-	

9. ПРАКТИЧНІ / СЕМІНАРСЬКІ ЗАНЯТТЯ, ЛАБОРАТОРНІ РОБОТИ

№ з/п	Назва лабораторної роботи	Кількість годин	
		Денна	Заочна
1	Експериментальне дослідження ефективності автоматизованих засобів обробки природомовних текстів різних мов.	4	-
2	Морфологічно-синтаксичний аналіз тексту на основі математичних методів засобами Python.	4	-
3	Застосування контекстно-вільної та контекстно-залежної граматики.	4	-

4	Застосування онтологічного підходу в автоматизованому семантичному аналізі.	4	-
5	Реалізація пошукової системи по тексту природньої мови.	4	-
6	Використання існуючих та розробка елементарних засобів машинного перекладу.	4	-
7	Розробка системи автоматизованого реферування тексту природньої мови.	4	-
8	Проектування системи природо мовного діалогу.	4	-

10. САМОСТІЙНА РОБОТА

Поглиблене опрацювання розглянутих на лекціях та розгляд суміжних тем.

№ з/п	Назва теми	Кількість годин	
		Денна	Заочна
1	Технології обробки природо мовної інформації в інформаційній безпеці.	6	-
2	Лямбда-числення в обробці природо мовної інформації.	8	-
3	Схожі та відмінні риси формальної граматики мови програмування та природньої мови.	10	-
4	Види семантик та способи їх моделювання.	14	-
5	Основи теорії інтертекстуальності.	8	-
6	Системи смислового аналізу Text Mining в машинному перекладі.	8	-
7	Лематизація як етап автоматизованого реферування.	10	-
8	Словникове забезпечення діалогових систем.	8	-
Разом		72	-

МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ

1. Методичні рекомендації до самостійної роботи з дисципліни «Технології обробки природо мовної інформації» для здобувачів освітнього ступеня бакалавр за спеціальністю 126 «Інформаційні системи та технології» усіх форм навчання (<https://er.chdtu.edu.ua/handle/ChSTU/1007>)

11. СИСТЕМА ОЦІНЮВАННЯ НАВЧАЛЬНИХ ДОСЯГНЕНЬ

11.1 МЕТОДИ КОНТРОЛЮ

Поточний контроль: усний контроль (в ході опитування під час захисту лабораторних робіт); письмовий контроль (при написанні контрольних робіт після засвоєння відповідних тем – 1-4 та 5-8); лабораторний контроль; спостереження за ходом виконання лабораторних робіт.

Модульний контроль: у вигляді письмової контрольної роботи здійснюється в кінці кожного модуля (після засвоєння 4-ї та 8-ї теми). В контрольній роботі міститься 8 запитань, відповіді на які у розгорнутому вигляді оцінюються від 1 до 2 балів за кожну повну відповідь (в залежності від складності запитання та об'єму очікуваної відповіді, загальною сумою 12 балів). Усі питання наведено в розділі «контрольні запитання» лабораторних робіт 1-4 для 1-го модуля та 5-8 для другого модуля (проводиться на 9-10 та 17-18 навчальному тижні за розкладом навчальних занять).

Рейтинговий контроль базується на результатах модульного контролю. Результати рейтингового контролю заносяться до журналу відображення успішності студентів (журнал рейтингового контролю), який зберігається в деканаті.

Семестровий контроль з дисципліни проводиться відповідно до навчального плану у вигляді семестрового іспиту (перелік питань міститься в підрозділі 11.2 силабусу). Здобувач вважається допущеним до семестрового контролю, якщо він виконав усі види робіт, передбачені навчальним планом на семестр.

Оцінка навчальних досягнень здобувачів вищої освіти за всіма видами контролю здійснюється за національною системою та ECTS:

Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90-100	A	відмінно	зараховано
82-89	B	добре	
74-81	C		
64-73	D	задовільно	
60-63	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання

0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни
------	---	--	---

11.2 ПИТАННЯ ДО ЗАЛІКУ

1. Охарактеризуйте загальні принципи мовної компетенції комп'ютерів.
2. Зазначте сфери діяльності, де застосовуються методи машинної обробки природньої мови.
3. Наведіть найбільш популярні інструментальні програмні засоби обробки природо мовної інформації.
4. На чому ґрунтується вибір тих чи інших інструментальних засобів обробки природо мовної інформації?
5. Опишіть спільні та відмінні риси обробки тексту для різних мов.
6. Охарактеризуйте відомі статистичні моделі природніх мов.
7. Яким чином проводиться дискретний аналіз природньої мови?
8. Які ви знаєте види неklasичних логік? Які з них можливо застосовувати при обробці тексту природньої мови?
9. Опишіть методику автоматизованого морфологічного та синтаксичного аналізу.
10. Перелічіть найбільш популярні синтаксичні аналізатори, звертаючи увагу на мову, з якою вони працюють та методи, що лежать в їх основі.
11. Яким чином можливо автоматизувати процес дослідження семантичного значення лексичної одиниці?
12. Наведіть інструментальні засоби, мови програмування Python, які спрощують обробку природньої мови.
13. Яким чином використовується функціональне програмування в дослідженні тексту?
14. Що являє собою формальна мова та як вона створюється?
15. Опишіть особливості та сферу застосування контекстно-вільної граматики.
16. Опишіть особливості та сферу застосування контекстно-залежної граматики.
17. Перелічіть різновиди генеративної граматики.
18. В чому полягають основні особливості трансформаційної граматики?
19. Опишіть базові аспекти граматики зв'язків.
20. Що саме являє собою модель «Смисл-Текст»?

21. В чому особливість онтологічного підходу в обробці інформації природньої мови?
22. Наведіть методи формальної семантики.
23. Опишіть базові правила семантики Монтегю.
24. Що таке дискурсний аналіз та яким чином можливо його автоматизувати?
25. Яким чином організований пошук у тексті.
26. Наведіть основні алгоритми текстового пошуку та їх базові особливості.
27. Зазначте особливості семантичного пошуку.
28. Опишіть сучасні інструментальні засоби автоматизованого перекладу.
29. В чому суть статистичного та гібридного методу в проектуванні та розробці систем автоматизованого перекладу.
30. Класифікуйте програмні засоби машинного перекладу та наведіть представників кожного класу.
31. Перелічіть сучасні автоматизовані системи реферування тексту.
32. Опишіть методи та алгоритми, що лежать в основі сучасних автоматизованих систем реферування тексту.
33. Наведіть етапи реферування текстової інформації.
34. Що означає термін інформаційна ентропія відносно процесу обробки природньої мови?
35. Що таке межа семантичного стиснення та яким чином вона визначається?
36. Поясніть суть систем природо мовного діалогу.
37. Що таке корпус текстів? Яким чином це поняття застосовується в проектуванні систем природо мовного діалогу?
38. Наведіть типи моделей природо мовного діалогу.
39. Зазначте сферу застосування систем природо мовного діалогу.
40. В чому полягає суть лінгво-комбінаторного моделювання погано формалізованих систем?

11.3 КРИТЕРІЇ ОЦІНЮВАННЯ

ДЕННА ФОРМА

Модуль	Критерії оцінювання знань	Кількість балів максимум
Змістовий модуль № 1		

Тема 1. Базові принципи комп'ютерної обробки природної мови.	6
Тема 2. Математичні методи аналізу природної мови.	7
Тема 3. Формальна граматики природної мови.	7
Тема 4. Формальна семантика природної мови.	10
<i>Всього за змістовим модулем №1</i>	30
Змістовий модуль № 2	
Тема 5. Алгоритми пошуку в тексті.	6
Тема 6. Проектування, розробка та використання засобів реалізації машинного перекладу.	6
Тема 7. Автоматизоване реферування текстової інформації.	10
Тема 8. Системи природо мовного діалогу.	8
<i>Всього за змістовим модулем №2</i>	30
Іспит	40
Разом	100

12. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Бенгфорт Б., Бильбро Р., Охеда Т. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. СПб.: Питер, 2019. 368 с.
2. Тарасенко Я.В. Метод семантичного стиснення текстової інформації для протидії комп'ютерній лінгвістичній стеганографії : дис. ... канд. технічних наук : 05.13.05. Черкаси, 2018. 188 с.
3. Bird S., Klein E., Loper E. Natural Language Processing with Python: Analyzing Text with the Natural language Toolkit. Sebastopol : O'Reilly Media, 2009. 504 p.

Допоміжна

4. Герасимова И.А. Формальная грамматика и интенциональная логика. М.: ИФ РАН, 2000. 156 с.
5. Підласий А.І. Інтелектуальні системи обробки неструктурованої інформації. Модуль 1. Інтелектуальні системи гуманітарного призначення: навчальний посібник. Черкаси: ЧДТУ, 2005. 171 с.
6. Руденко О.Г., Бодянський Є.В. Штучні нейронні мережі. Х.: Компанія СМІТ, 2006. 404 с.
7. Снитюк В.Є. Прогнозування. Моделі. Методи. Алгоритми: навчальний посібник. К.: Маклаут, 2008. 364 с.
8. Гнатієнко Г.М., Снитюк В.Є. Експертні технології прийняття рішень: монографія. К.: Маклаут, 2008. 444 с.

13. ІНФОРМАЦІЙНІ РЕСУРСИ

1. <https://www.lektorium.tv/course/22761>
2. https://www.youtube.com/playlist?list=PL3FW7Lu3i5Jsnh1rnUwq_TcylNr7EkRe6
3. <https://er.chdtu.edu.ua/simple-search?filterquery=тарасенко%2c+ярослав+володимирович&filtername=author&filtertype>equals>

14. ПЕРЕЛІК НОРМАТИВНИХ ДОКУМЕНТІВ

1. Положення про організацію контролю та оцінювання якості навчання студентів (<https://chdtu.edu.ua/normative/regulations/item/420-polozhennya-pro-organizatsiyu-kontrolyu-ta-otsinyuvannya-yakosti-navchannya-studentiv>).
2. Положення про організацію освітнього процесу в Черкаському державному технологічному університеті (<https://chdtu.edu.ua/normative/regulations/item/3636-polozhennya-pro-orhanizatsiyu-osvitnoho-protsesu-v-cherkaskomu-derzhavnomu-tekhnologichnomu-universyteti>).
3. Кодекс академічної доброчесності Черкаського державного технологічного університету (<https://chdtu.edu.ua/normative/regulations/item/8892-kodeks-akademichnoyi-dobrochesnosti-cherkaskoho-derzhavnoho-tekhnologichnoho-universytetu-zizminamy>).

15 ПОЛІТИКА ДИСЦИПЛІНИ

1.	Постійна обов'язкова частина	
1.1	Контрольна робота №1 (тема 1, 2, 3, 4)	10
1.2	Контрольна робота №2 (тема 5, 6, 7, 8)	10
1.3	Лабораторна робота №1	5
1.4	Лабораторна робота №2	5
1.5	Лабораторна робота №3	5
1.6	Лабораторна робота №4	5
1.7	Лабораторна робота №5	5
1.8	Лабораторна робота №6	5
1.9	Лабораторна робота №7	5
1.10	Лабораторна робота №8	5
	Разом	60
2.	Додаткова варіативна частина	
2.1	Активна робота на занятті	5
2.2	Виступ із доповіддю на науковому семінарі	25
2.3	Призове місце на конкурсі студентських робіт ЧДТУ	15

	Разом	45
3.	Штрафна частина	
3.1	Пропуск одного заняття.	-1
3.2	Захист лабораторної роботи №1 після 3-го навчального тижня	-2
3.3	Захист лабораторної роботи №2 після 5-го навчального тижня	-2
3.4	Захист лабораторної роботи №3 після 7-го навчального тижня	-2
3.5	Захист лабораторної роботи №4 після 9-го навчального тижня	-2
3.6	Захист лабораторної роботи №5 після 11-го навчального тижня	-2
3.7	Захист лабораторної роботи №6 після 13-го навчального тижня	-2
3.8	Захист лабораторної роботи №7 після 15-го навчального тижня	-2