КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ
КАФЕДРА МАТЕМАТИЧНОЇ ІНФОРМАТИКИ

«ЗАТВЕРДЖУЮ»
Заступник декана
з навчальної роботи
_____ Олена КАШПУР
«31» серпня 2022 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
МЕТОДИ УЩІЛЬНЕННЯ ТА ПОШУКУ ДАНИХ

для студентів

галузь знань	**12 "Інформаційні технології"**
спеціальність	**122 "Комп'ютерні науки"**
освітній рівень	**бакалавр**
освітня програма	**Інформатика**
вид дисципліни	**вибіркова**

Форма навчання	**стаціонар**
Навчальний рік	**2021/2022**
Семестр	**4**
Кількість кредитів ECTS	**4**
Мова викладання, навчання та оцінювання	**українська**
Форма заключного контролю	**залік**

Викладач: д.ф.-м.н., доцент Завадський І.О

КИЇВ – 2022

Розробник: Завадський І.О., д.ф.-м.н., доцент кафедри математичної інформатики, кафедра математичної інформатики

ЗАТВЕРДЖЕНО
Зав. кафедри «Математичної інформатики»
_____ Василь ТЕРЕЩЕНКО

Протокол № 1 від "29" серпня 2022 р.

Схвалено гарантом освітньо-професійної програми «Інформатика»

_____ Людмила ОМЕЛЬЧУК «31» серпня 2022 рік
(підпис) (прізвище та ініціали)

Схвалено науково-методичною комісією факультету комп'ютерних наук та кібернетики

Протокол від «31» серпня 2022 року № 1
Голова науково-методичної комісії _____ Людмила ОМЕЛЬЧУК

1. Мета дисципліни "Методи ущільнення та пошуку даних" – ознайомлення з основними досягненнями в кількох найбільш актуальних галузях комп'ютерної науки, пов'язаних із кодуванням, пошуком та обробкою інформації.

2. Попередні вимоги до опанування або вибору навчальної дисципліни:

1. Знати: основи з дисциплін "Програмування", "Дискретна математика", "Лінійна алгебра та аналітична геометрія", "Математичний аналіз".
2. Вміти: проводити аналіз задач, визначати оцінки складності алгоритмів їх розв'язання; застосовувати алгоритми на практиці та вимірювати їхню ефективність.
3. Володіти методами побудови та оцінювання алгоритмів як засобом розв'язання обчислювальних задач, що постають у різних галузях кібернетики.

3. Анотація навчальної дисципліни:

Навчальна дисципліна «Методи ущільнення та пошуку даних» є складовою освітньо-професійної програми «Інформатика» підготовки фахівців за освітньо-кваліфікаційним рівнем «бакалавр» галузі знань 12 «Інформаційні технології» спеціальності 122 «Комп'ютерні науки». Вона є навчальною дисципліною, що пропонується студенту на вибір і входить до вибіркового переліку №2. Викладається у 4 семестрі 2 курсу бакалаврату в обсязі *4 кредити ECTS.*

Курс складається з 2 *змістових частини*. Протягом його вивчення передбачено виконання 2 *контрольних робіт та* 2 *електронних тестувань*. Робота студента протягом семестру оцінюється у формі заліку.

4. Завдання (навчальні цілі): сформувати знання класичних і сучасних методів і підходів до розв'язання фундаментальних задач кібернетики, вміння застосовувати ці методи й підходи на практиці та розвинути алгоритмічне мислення на рівні, достатньому для самостійного оцінювання, вдосконалення та винайдення алгоритмів розв'язання різноманітних задач в галузі обробки даних.

5. Результати навчання за дисципліною:

Результат навчання (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність)		**Форми (та/або методи і технології) викладання і навчання**	**Методи оцінювання та пороговий критерій оцінювання (за необхідності)**	**Відсоток у підсумковій оцінці з дисципліни**
Код	**Результат навчання**			
РН1.1	Знати та розуміти поняття інформаційної надлишковості та ентропії, основні принципи стискання даних без інформаційних втрат.	Лекції, самостійна робота	Контрольна робота 1, тестування 1 (60% правильних відповідей)	3%
РН 1.2	Знати правила побудови множин кодових слів, а також методи кодування й декодування таких стискальних кодів, як коди Хафмана, арифметичні коди, коди Лемпеля-Зіва, Голомба, Елайеса, Фібначчі, коди, що базуються на асиметричних системах числення.	Лекції, практичні заняття, самостійна робота		15%
РН 1.3	Знати та розуміти поняття однозначної декодовності, універсальності, повноти та асимптотичної щільності коду.	Лекція, самостійна робота		3%

РН 1.4	Знати та розуміти основні методики розв’язання задачі пошуку рядка в тексті.	Лекції, практичні заняття, самостійна робота	Контрольна робота 2, тестування 2 (60% правильних відповідей)	4%
РН 1.5	Знати такі алгоритми точного пошуку рядка в тексті, як KMP, BMH, QS, Shift-And/Or, TVSBS, BR, MAW, BOM/EBOM, SBNDM, Z/RZ, Hashq.	Лекції, практичні заняття, самостійна робота		15%
РН 2.1	Вміти визначати та порівнювати теоретичні характеристики джерела даних та методів їхнього стиснення, зокрема визначати однозначну декодовність, повноту та універсальність коду.	Лекції, практичні заняття, самостійна робота	Контрольна робота 1 (60% правильних відповідей), домашнє завдання	10%
РН 2.2	Вміти застосовувати алгоритми стиснення даних на прикладах різнотипних інформаційних повідомлень.	Лекції, практичні заняття, самостійна робота	Контрольна робота 2, тестування 2 (60% правильних відповідей), домашнє завдання	15%
РН 2.3	Вміти застосовувати алгоритми пошуку рядка в тексті.	Лекції, практичні заняття, самостійна робота	Контрольна робота 2, тестування 2 (60% правильних відповідей), домашнє завдання	15%
РН 3.1	Обґрунтовувати вибір оптимального методу стиснення без втрат залежно від характеристик даних, що стискаються.	Лекції, практичні заняття, самостійна робота	Контрольна робота 1, тестування 1 (60% правильних відповідей)	5%
РН 3.2	Обґрунтовувати вибір оптимального методу пошуку рядка в тексті залежно від параметрів тексту та шуканого рядка.	Лекції, практичні заняття, самостійна робота	Контрольна робота 2, тестування 2 (60% правильних відповідей)	5%
РН 4.1	Організовувати власну самостійну роботу для досягнення результату.	Лекція, самостійна робота	Контрольні роботи 1 і 2, тестування 1 і 2, домашнє завдання	5%
РН 4.2	Відповідально ставитися до виконуваних робіт, нести відповідальність за їхню якість.	Лекція, самостійна робота		5%

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

Результати навчання дисципліни / Програмні результати навчання	РН 1.1	РН 1.2	РН 1.3	РН 1.4	РН 1.5	РН 2.1	РН 2.2	РН 2.3	РН 3.1	РН 3.2	РН 4.1	РН 4.2
ПРН4. Проектувати, розробляти та аналізувати алгоритми розв'язання обчислювальних та логічних задач, оцінювати ефективність та складність алгоритмів на основі застосування формальних моделей алгоритмів та обчислюваних функцій.	+	+	+	+	+	+	+	+	+	+	+	+

7. Схема формування оцінки.

7.1. Форми оцінювання студентів:

- семестрове оцінювання:

1. Контрольна робота 1 (письмова робота): РН 1.1, РН 1.2, РН 1.3, РН 2.1, РН 2.2, РН 3.1, РН 4.1, РН 4.2 — 30 балів / 18 балів.

2. Контрольна робота 2 (письмова робота): РН 1.4, РН 1.5, РН 2.3, РН 3.2, РН 4.1, РН 4.2 — 30 балів / 18 балів.

3. Домашні завдання (письмова робота): РН2.1, РН 2.2, РН 2.3, РН 4.1, РН 4.2 — 10 балів / 6 балів.

5. Тестування 1 (електронний тест): РН 1.1, РН 1.2, РН 1.3, РН 2.1, РН 2.2, РН 3.1, РН 4.1, РН 4.2 — 15 балів / 9 балів.

6. Тестування 2 (електронний тест): РН 1.4, РН 1.5, РН 2.3, РН 3.2, РН 4.1, РН 4.2 — 15 балів / 9 балів.

- підсумкове оцінювання (у формі заліку):

- залікові бали визначаються як сума оцінок/балів за всіма успішно оціненими результатами навчання, передбаченими даною програмою;
- оцінки нижче від мінімального порогового рівня не додаються;
- мінімальний пороговий рівень для сумарної оцінки за всіма компонентами становить 60% від максимально можливої кількості балів.

7.2. Організація оцінювання:

Обов'язковим є виконання завдань, винесених на самостійну роботу, контрольних робіт і тестувань за графіком робочої програми.

Терміни проведення форм оцінювання:

1. Контрольна робота 1: до 8 тижня семестру.

2. Контрольна робота 2: до 15 тижня семестру.

4. Домашнє завдання 1: до 5 тижня семестру.

5. Домашнє завдання 2: до 10 тижня семестру.

6. Тестування 1: до 4 тижня семестру.

7. Тестування 2: до 12 тижня семестру.

Студент має право на однократне перескладання кожної контрольної роботи із можливістю отримання максимально 90% початково визначених за цю контрольну роботу балів. Термін перескладання визначається викладачем.

Тематика робіт, що виконуються студентами.

Контрольна робота 1 містить 5 теоретичних завдань, що перевіряють розуміння сутності технології стискання даних без втрат інформації та її обмежень, а також 2 практичних завдання на застосування певних стискальних кодів на конкретних прикладах.

Контрольна робота 2 містить 2 практичних завдання на застосування алгоритмів пошуку до конкретних даних.

Домашнє завдання 1 полягає у оцінці ефективності певних кодів та пошуку оптимальних способів кодування джерел даних із заданими характеристиками.

Домашнє завдання 2 полягає в інтерпретації програмного коду пошукових алгоритмів.

Тестування 1 складається з 10 завдань з тематики першої частини курсу.

Тестування 2 складається з 6 завдань з тематики другої частини курсу.

6.3. Шкала відповідності оцінок

Зараховано	60-100
Не зараховано	0-59

.

8. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
ТЕМАТИЧНИЙ ПЛАН ЛЕКЦІЙ І СЕМІНАРСЬКИХ ЗАНЯТЬ

№ лекції	Назва лекції	Кількість годин		
		Лекції	Практичні	С/Р
Частина 1. Алгоритми стискального кодування				
1	**Тема 1.** Поняття ентропії інформаційного повідомлення, постановка та різновиди задач стискального кодування. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 1.*	4		4
2	**Тема 2.** Класичні методи кодування без втрат, коди Хафмана та адаптивні коди Хафмана. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 1*	2	2	6
3	**Тема 3.** Арифметичне кодування. Метод FSE. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 1*	2	2	10
4	**Тема 4.** Кодування Лемпеля-Зіва та Лемпеля-Зіва-Велча, їх застосування. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 1*	2	2	4

5	**Тема 5.** Універсальність, повнота та асимптотична щільність кодів. Нерівність Крафта. Коди Елайеса. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 1*	2	2	4
6	**Тема 6.** Коди Голомба та коди Фібоначчі. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 2*	2	2	4
7	**Тема 7.** Мультироздільникові та реверсні мультироздільникові коди. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 2*	2	2	4
Контрольна робота 1		2		
Частина 2. Пошукові алгоритми				
8	**Тема 8.** Постановка та різновиди задач пошуку рядка в тексті. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 2*	4	1	2
9	**Тема 9.** Алгоритми префіксного пошуку. Алгоритм Кнута-Моріса-Пратта, Shift-And, Shift-Or. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 2*	2	1	4
10	**Тема 10.** Алгоритми суфіксного пошуку. Алгоритм Боєра-Мура та його похідні. *Самостійна робота: опрацювання лекційного матеріалу. Виконання домашнього завдання 2*	2	2	4
11	**Тема 11.** Пошук на основі q-грам. Алгоритми Бері-Равіндрана, TVSBS, MAW, Hashq. *Самостійна робота: опрацювання лекційного матеріалу.*	2	1	4
12	**Тема 12.** Пошук на основі неповних q-грам. Алгоритми Z/RZ. *Самостійна робота: опрацювання лекційного матеріалу.*	2	1	4
13	**Тема 13.** Пошук з використанням детермінованих скінченних автоматів. Алгоритми BDM, BOM та EBOM. *Самостійна робота: опрацювання лекційного матеріалу.*	2	1	4
14	**Тема 14.** Пошук з використанням недетермінованих скінченних автоматів. Алгоритми BNDM, SBNDM, FSBNDM. *Самостійна робота: опрацювання лекційного матеріалу.*	2	1	4
Контрольна робота 3		2		
ВСЬОГО		36	20	62

Загальний обсяг ***120 годин***, в тому числі:
Лекції – ***36 годин.***
Практичні – ***20 годин.***
Самостійна робота – ***62 годин.***

9. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна:

1. G.Navarro. M. Raffinot. Flexible pattern matching in strings: Cambridge University press, 2002. – 224 p.
2. S. Faro and T. Lecroq: The exact online string matching problem: a review of the most recent results. ACM Computing Surveys (CSUR), 45(2) 2013, p. article 13.
3. Ахо Х., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов. М.: Мир, 1979. – 536.
4. Кнут Д. Искусство программирования для ЭВМ , т.2 – М., 1977, 720с.
5. D. Salomon: Variable-Length Codes for Data Compression, Springer-Verlag, London, U.K., 2007, 196 p.

Додаткова:

1. D.E. Knuth, Jr. J.H. Morris, V.R. Pratt, 1977. Fast pattern matching in strings. SIAM J. Comput. 6, 1, 323–350.
2. S. Wu and U. Manber: Fast text searching allowing errors. Commun. ACM, 35(10), 1992, pp. 83–91.
3. R. Baeza-Yates, G.H. Gonnet. A new approach to text searching. Commun. ACM 35 (10), 1992, pp. 74–82.
4. T. Berry and S. Ravindran: A fast string matching algorithm and experimental results, in Proceedings of the Prague Stringology Club Workshop, Edited by Jan Holub and Milan Simanek, Eds. Czech Technical University in Prague, Czech Republic, 1999, pp. 16–28.
5. R. S. Boyer and J. S. Moore: A fast string searching algorithm. Commun. ACM, 20(10) 1977, pp. 762–772.
6. D. Cantone and S. Faro: Improved and self-tuned occurrence heuristics. Journal of Discrete Algorithms, 28 2014, pp. 73–84.
7. S. Faro and T. Lecroq: Efficient variants of the backward-oracle-matching algorithm, in Proceedings of the Prague Stringology Conference, J. Holub and J. Zdarek, Eds. Czech Technical University in Prague, Czech Republic, 2008, pp. 146–160.
8. D. Huffman: A method for the construction of minimum-redundancy codes. Proc. IRE, 40 1952, pp. 1098–1101.
9. S. T. Klein and M. Ben-Nissan: On the usefulness of Fibonacci compression codes. Computer Journal, 53(6) 2010, pp. 701–716.
10. N. Brisaboa, A. Farina, G. Navarro, and M. Esteller: (s,c)-dense coding: an optimized compression code for natural language text databases, in Proc. Symposium on String Processing and Information Retrieval SPIRE'03, no. 2857 in Lecture Notes in Computer Science, Manaus, Brazil, 2003, Springer-Verlag, Berlin, pp. 122–136.
11. A.V. Anisimov, I.O. Zavadskyi. Variable-Length Prefix Codes With Multiple Delimiters // IEEE Transactions on Information Theory, vol. 63, issue 5, p. 2885-2895. –2019.
12. I.O. Zavadskyi. Fast exact pattern matching in a bitstream and 256-ary strings. Proceedings of the Prague Stringology Conference. – 2020, pp. 33–47.