

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ
Кафедра математичної інформатики



«ЗАТВЕРДЖУЮ»

Заступник декана
з навчальної роботи

Кашпур О.Ф.

серпня 2020 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Розподілені системи обробки інформації/Distributed
Systems of Information Processing
для студентів

галузь знань	12 "Інформаційні технології"/"Information Technologies"
спеціальність	122 "Комп'ютерні науки"/"Computer Science"
освітній рівень	магістр/masters
освітня програма	Математичні методи штучного інтелекту/ Mathematical methods of
Artificial Intelligent	
вид дисципліни	Обов'язкова навчальна дисципліна/ mandatory

Форма навчання	денна
Навчальний рік	2020/2021
Семестр	2
Кількість кредитів ECTS	4
Мова викладання, навчання та оцінювання	англійська, українська\
English, Ukrainian	
Форма заключного контролю	іспит/exam

Викладачі: професор Глибовець Микола Миколайович, д.ф.-м.н.

Пролонговано: на 20²¹/20²² н.р. () « 11 » 02 20²¹р.
на 20 /20 н.р. () « » 20 р.

КИЇВ – 2020

Розробник: Глибовець Микола Миколайович, д. ф.-м. н., проф.

ЗАТВЕРДЖЕНО

Завідувач кафедри математичної інформатики

_____ Терещенко В.М.
(підпис)

Протокол № 3 від «21» травня 2020р.

Схвалено Гарантом освітньо-наукової програми «Математичні методи штучного

інтелекту» _____ (Завадський І.О.)
(підпис)

«28» серпня 2020р.

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

Протокол від «28» серпня 2020року № 1
Голова науково-методичної комісії _____ (Омельчук Л.Л.)
(підпис)

«28» серпня 2020року

1. Мета дисципліни — дати сучасні знання про теоретичні основи, технології побудови розподілених систем обробки інформації, зосередитися на проблемах та шляхах їх подолання при роботі з великими даними.

/

Discipline aim. The aim of the discipline is to provide up-to-date knowledge of theoretical foundations, technologies of construction distributed systems, put attention on problems and ways to overcome them when working with big data.

2. Попередні вимоги до опанування або вибору навчальної дисципліни/

Preliminary demands to master or choice of the course discipline:

1. *Знати:* базову дисципліну «Кластеризація та класифікація інформації», теорію алгоритмів, основні структури даних, реляційні та нереляційні бази даних, мови програмування. Знання технічної англійської мови на рівні B1.
2. *Вміти:* розробляти, аналізувати та застосовувати програмні системи для розв'язання завдань та прикладних задач, використовуючи сучасні методи розробки програм.

/

1. To know: the basic discipline "Clusterization and classification of information", theory of algorithms, basic data structures, relational and non-relational databases, programming languages. Level B1 technical English skills.
2. To be able to: develop, analyze, and apply software systems to solve problems and applied tasks using modern software development methods.

3. Анотація навчальної дисципліни / Synopsis of the course:

Дисципліна «Розподілені системи обробки інформації» належить до переліку обов'язкових навчальних дисциплін освітньо-наукової програми підготовки фахівців «Математичні методи штучного інтелекту» за другим (магістерським) рівнем вищої освіти галузі знань 12 «Інформаційні технології» спеціальності 122 «Комп'ютерні науки». Вона забезпечує професійний розвиток магістра, спрямована на формування теоретичних основ, технології побудови та використання розподілених систем, зокрема, розуміння проблематики роботи з великими даними в розподілених застосуваннях.

Дана дисципліна є обов'язковою навчальною дисципліною за програмою «Математичні методи штучного інтелекту». Викладається у 2 семестрі 1 курсу магістратури в обсязі – 4 кредитів ECTS.

У курсі передбачено 3 змістових частини, 2 контрольні роботи, 2 лабораторні роботи. Завершується дисципліна екзаменом в 2 семестрі 1 курсу магістратури.

/

The discipline "Distributed Systems of Information Processing" belongs to the list of mandatory disciplines as a component of the educational-scientific training program for the second (master's) level of higher education in the field of knowledge 12 "Information Technology" specialty 122 "Computer Science", educational-scientific program "Mathematical methods of Artificial Intelligence". It provides the professional development of master students, aimed at the formation of theoretical foundations, technology of construction and use of distributed systems in the area of work with BigData.

4. Завдання (навчальні цілі) / Objectives of study: Основними завданнями дисципліни «Розподілені системи обробки інформації» є набуття знань, умінь та навичок

(компетентностей) на рівні новітніх досягнень в області розподілених систем відповідно до освітньої кваліфікації магістр зі штучного інтелекту. Зокрема, розвивати:

- здатність застосовувати знання у практичних ситуаціях (ЗК2);
- здатність спілкуватися іноземною мовою (ЗК5);
- здатність до дослідження та аналізу надвеликих масивів даних із складною неоднорідною і/або невизначеною структурою для прийняття зважених бізнес-рішень (СК3);
- здатність застосовувати методи і засоби організації великих даних для проектування масштабованих інфраструктур консолідації ресурсів зберігання, дослідження, управління, захисту та обслуговування інформації, розв'язання завдань моделювання та прогнозування стратегічних напрямків розвитку бізнесу (СК4);
- здатність аналізувати сучасні світові тенденції розвитку комп'ютерних наук та уявляти перспективи розвитку інформаційних технологій, моделювати процеси розвитку і трансформації інформаційно-комунікаційних технологій в практичній професійній роботі (СК11);
- розуміння економічних переваг інноваційного розвитку ІТ підприємств (новітні підходи організації, застосування програмних, апаратних, мережних, математичних, технологічних, ергономічних та інших засобів) з метою вирішення актуальних задач підвищення конкурентоспроможності галузі; здатність розв'язувати складні задачі і проблеми проектування корпоративного інформаційного середовища, що передбачає здійснення інновацій (СК12);
- здатність проектування динамічних веб-додатків як інформаційної системи із застосуванням об'єктно-орієнтованих технологій програмування, зокрема сучасних програмних засобів підтримки взаємодії клієнта та сервера із застосуванням розподілених систем керування базами даних, супроводження та оптимізація веб-сторінок (СК14);

/

Objectives (learning objectives): acquiring knowledge, skills, and competencies at the level of the latest achievements in distributed systems. In particular, to develop:

- the ability to apply knowledge in practical situations;
- the ability to communicate in a foreign language;
- the ability to investigate and analyze large-scale datasets of complex heterogenous and/or undetermined structure in the aim of well-targeted decision making;
- the ability to apply Big Data methods and tools in order to design scalable infrastructures of storage consolidation, research, manage, secure and maintain data, solving problems of modelling and forecasting of strategic business development threads;
- the ability to analyze modern world trends in the development of computer science and imagine the prospects for the development of information technology, to model the processes of development and transformation of information and communication technologies in practical professional work;
- understanding of economic preferences of innovative development of IT enterprises (latest approaches to organization, application of software, hardware, network, mathematical, technological, ergonomic and other means) in order to solve urgent problems of increasing the competitiveness of the industry; ability to solve complex problems and problems of designing a corporate information environment that involves innovation;
- the ability to design the dynamic web-applications as information system implementing the object-oriented programming technologies, in particular modern program tools which support the client-server interaction with the application of distributed database management systems, web pages maintaining and optimization.

5. Результати навчання за дисципліною/ Results of learning:

Результат навчання (РН) (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)	Відсо- ток у підсум ковій оцінці з дисцип ліни
Код	Результат навчання			
РН 1.1	Знати основні проблеми при побудові розподілених систем / To know the main problems of development distributed systems	Лекція / Lecture	Активна робота на лекції, усні відповіді, Контрольна робота 1 (60% правильних відповідей), іспит / Active work on lectures, oral answers, Test 1 (60% correct answers), exam	20%
РН 1.2	Знати основні складові розподілених систем / To know the main parts of distributed systems	Лекція / Lecture		
РН 1.3	Знати основні підходи до побудови розподілених систем \ / To know the main approaches to distributed systems development	Лекція / Lecture	Активна робота на лекції, усні відповіді, Контрольна робота 2 (60% правильних відповідей), іспит / Active work on lectures, oral answers, Test 2 (60% correct answers), exam	15%
РН 1.4	Знати основні підходи до роботи з даними в розподілених системах / To know the basic approaches for working with data in distributed systems	Лекція / Lecture	Активна робота на лекції, усні відповіді, Контрольна робота 3 (60% правильних відповідей), іспит / Active work on lectures, oral answers, Test 3 (60% correct answers), exam	15%
РН 2.1	Вміти будувати Reliable, Scalable and Maintainable застосування / Be able to build Reliable, Scalable and Maintainable applications	Лабораторна робота, самостійна робота/ Laboratory work, individual work	Захист лабораторної роботи 1, Виконання завдань, винесених на самостійну роботу, іспит / laboratory work 1 defense, Tasks assigned to independent work, exam	15%
РН 2.2	Вміти розробляти та аналізувати розподілені системи, що працюють з великими даними / Be able to develop and analyze distributed systems that work with big data	Лабораторна робота, самостійна робота/ Laboratory work, individual work	Захист лабораторної роботи 2, Виконання завдань, винесених на самостійну роботу, іспит / laboratory work 2 defense, Tasks assigned to independent work, exam	15%
РН 2.3	Вміти організувати реплікацію та розбиття даних в розподілених системах та вміти будувати системи, що будуть це ефективно використовувати / Be able to make replication and partitioning of data in distributed systems and be able to build systems that will use it effectively	Лабораторна робота, самостійна робота/ Laboratory work, individual work	Захист лабораторної роботи 3, Виконання завдань, винесених на самостійну роботу, іспит / laboratory work 3 defense, Tasks assigned to independent work, exam	15%

PH 4.1	Відповідально ставитися до виконуваних робіт, нести відповідальність за їх якість / Responsibly treat the works performed, be responsible for their quality	Самостійна робота / Individual work	Виконання завдань, винесених на самостійну роботу / Accomplishment of tasks assigned to Individual work	5%
--------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------	---------------------------------------------------------------------------------------------------------------	----

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання / Correspondence between learning results and program study results

Результати навчання дисципліни/ Програмні результати навчання (з опису освітньої програми)	PH 1.1	PH 1.2	PH 1.3	PH 1.4	PH 2.1	PH 2.2	PH 2.3	PH 4.1
ПРН2. Використовувати моделі та методи прийняття рішень на основі теорії нечітких множин та в умовах невизначеності і ризиків в процесі управлінської діяльності за галузями. / Use models and methods of decision-making based on the theory of fuzzy sets and in conditions of uncertainty and risk in the process of management activities by industry.					+		+	
ПРН4. Аналізувати великі дані та моделювати високорівневі абстракції у великих наборах даних різної природи, проектувати сховища великих даних, для видобутку даних і знань, візуалізувати великі дані, будувати і оцінювати регресивні моделі, що генеруються на основі великих даних. / Analyze big data and model high-level abstractions in large data sets of different nature, design big data repositories to extract data and knowledge, visualize big data, build and evaluate regression models generated based on big data.						+	+	
ПРН5. Вирішувати складні проблеми, що вимагають систем з великою обчислювальною потужністю для забезпечення масштабованості паралельних алгоритмів і програм. / Solve complex problems that require systems with high computing power to ensure the scalability of parallel algorithms and programs.	+	+	+	+	+	+	+	+
ПРН6. Використовувати розподілені високопродуктивні обчислювальні технології для забезпечення ефективного вибору та використання консолідованих ресурсів і послуг. / Use distributed high-performance computing		+	+					+

technologies to ensure the efficient selection and use of consolidated resources and services.								
ПРН7. Вміти використовувати обчислювальні системи надвеликої потужності для виконання парадигми програмування мультипроцесорних обчислень, розробляти ефективні паралельні алгоритми складних виробничих задач, застосовувати хмарні платформи та їх віртуалізацію. / Be able to use ultra-high-power computing systems to perform the paradigm of programming multiprocessor computing, to develop effective parallel algorithms for complex production tasks, to use cloud platforms and their virtualization.								+

7. Схема формування оцінки.

7.1. Форми оцінювання студентів:

- семестрове оцінювання:

1. Активна робота на лекції, усні відповіді / *Active work on lectures, oral answers*: РН1.1, РН1.2, РН1.3, РН1.4 — 5 балів/3 бали;
2. Виконання завдань, винесених на самостійну роботу / *Tasks assigned to independent work*: РН2.1, РН2.2, РН2.3, РН4.1 — 5 балів/3 бали;
3. Контрольна робота 1: РН1.1, РН1.2 — 10 балів/6 балів;
4. Контрольна робота 2: РН1.3 — 10 балів/6 балів;
5. Контрольна робота 3: РН1.4 — 5 балів/3 балів;
6. Захист лабораторної роботи 1: РН 2.1 — 5 балів/3 бали;
7. Захист лабораторної роботи 2: РН 2.2 — 10 балів/6 балів;
8. Захист лабораторної роботи 3: РН 2.3 — 10 балів/6 балів;

- підсумкове оцінювання (у формі іспиту):

- максимальна кількість балів, які можуть бути отримані студентом: 40 балів;
- результати навчання, які будуть оцінюватись: РН1.1, РН1.2, РН1.3, РН1.4, РН2.1, РН2.2, РН2.3;
- форма проведення і види завдань: письмова робота.
Види завдань: 8 тестових та 4 письмових завдання.

- final assessment (in the form of an exam):

- the maximum number of points that can be obtained by a student: 40 points;
- learning outcomes that will be evaluated: РН1.1, РН1.2, РН1.3, РН1.4, РН2.1, РН2.2, РН2.3;
- form and types of tasks: written work.
Types of tasks: 8 test and 4 written tasks.

7.2. Організація оцінювання / Evaluation process:

Обов'язковим є виконання завдань, винесених на самостійну роботу, лабораторних робіт та модульних контрольних робіт за графіком робочої програми./ It is mandatory to perform tasks assigned to independent work, laboratory work and modular tests according to the schedule of the work program.

Терміни проведення форм оцінювання / Deadlines:

1. Контрольна робота 1 (тест): до 6 тижня семестру. Test1: up to the end of 6 weeks of the semester.

2. Контрольна робота 2 (тест): до 10 тижня семестру. Test2: up to the end of 10 weeks of the semester.

3. Контрольна робота 3 (тест): до 14 тижня семестру. Test2: up to the end of 14 weeks of the semester.

3. Лабораторна робота 1: до 4 тижня семестру. Laboratory work 1: up to the end of 4 weeks of the semester.

4. Лабораторна робота 2: до 9 тижня семестру. Laboratory work 2: up to the end of 9 weeks of the semester.

5. Лабораторна робота 3: до 14 тижня семестру. Laboratory work 3: up to the end of 14 weeks of the semester.

Студент має право на одне перескладання кожної контрольної роботи із можливістю отримання максимально 80% початково визначених за цю контрольну роботу балів. Термін перескладання визначається викладачем.

У випадку відсутності студента з поважних причин відпрацювання та перездачі контрольних робіт здійснюються у відповідності до „Положення про порядок оцінювання знань студентів при кредитно-модульній системі організації навчального процесу” від 1 жовтня 2010 року.

У разі неякісного виконання лабораторної роботи, викладач має право не зарахувати лабораторну роботу, або знизити за неї бали.

Студент має право здавати лабораторні роботи після закінчення визначеного для них терміну, але з втратою одного балу за кожен тиждень, який пройшов з моменту закінчення терміну її здачі.

The student has the right to one retake for each test with the possibility of obtaining a maximum of 80% of the points initially determined for this test. The term of retake is determined by the teacher.

In case of absence of the student for valid reasons working off and retaking of control works are carried out according to "Regulations on the order of an estimation of knowledge of students at the credit-modular system of the organization of educational process" from October 1, 2010.

In case of poor performance of laboratory work, the teacher has the right not to enroll in laboratory work, or to reduce scores for it.

The student has the right to take laboratory works after the expiration of the term set for them, but with the loss of one point for each week that has passed since the expiration of its submission.

7.3 Шкала відповідності оцінок / Mark correspondence scale

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59

8. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ ТЕМАТИЧНИЙ ПЛАН ЛЕКЦІЙ І ЛАБОРАТОРНИХ ЗАНЯТЬ

№	Назва лекції	Кількість годин
---	--------------	-----------------

лек-ції		Лекції	Семінари/ лабораторні	Самостійна робота
Частина I. Основи розподілених систем Part I. "Foundations of distributed systems"				
1	<p>Тема 1. Основи побудови надійних та масштабованих застосувань</p> <p><i>Самостійна робота:</i> Провести аналіз програми і змісту курсу та рекомендованої літератури.</p> <p>/</p> <p>Theme 1. Reliable, Scalable, and Maintainable Applications.</p> <p><i>Individual work:</i> To analyze the program and course content, as well as recommended literature.</p>	2		8
2	<p>Тема 2. Моделі даних та мови запитів</p> <p><i>Самостійна робота:</i> Побудувати систему, що використовує одночасно реляційну і не реляційну моделі даних</p> <p>/</p> <p>Theme 2. Data Models and Query Languages.</p> <p><i>Individual work:</i> Develop system that used relational and NoSQL models</p>	2		10
3	<p>Тема 3. Зберігання та пошук інформації.</p> <p><i>Самостійна робота:</i> Порівняти стовбчикові та рядкові сховища даних</p> <p>/</p> <p>Theme 3. Storage and Retrieval.</p> <p><i>Individual work:</i> Compare table and row models of data storages</p>	2		10
4	<p>Тема 4. Проблеми кодування інформації</p> <p><i>Самостійна робота:</i> Використати Apache Thrift або gRPC для обміну даними в розподіленій системі</p> <p>/</p> <p>Theme 4. Encoding and Evolution.</p> <p><i>Individual work:</i> Build distributed system that use Apache Thrift or gRPC for data exchange</p>	2	2	10
5	<p>Тема 5. Генетичні алгоритми в задачах комбінаторної оптимізації</p> <p><i>Самостійна робота:</i> Для обраної задачі побудувати розв'язок з використанням генетичних алгоритмів</p> <p>/</p> <p>Theme 5. Genetic algorithms in combinatorial optimization problems</p> <p><i>Individual work:</i> To construct a solution using genetic algorithms for the chosen task</p>	2		5

<i>Контрольна робота 1/ Control work 1</i>		1		
Частина II. Розподілені дані/ Part II. Distributed Data				
6	Тема 6. Реплікація <i>Самостійна робота:</i> Побудувати застосування, що використовує реплікацію master-master. / Theme 6. Replication. <i>Individual work:</i> Develop application that used master-master replication	2		5
7	Тема 7. Розбиття даних в розподілених системах <i>Самостійна робота:</i> Побудувати застосування, що буде використовувати розбиття даних на декілька сховищ. / Theme 7. Partitioning. <i>Individual work:</i> Develop application that will use sharding of data	2		7
8	Тема 8. Розподілені транзакції <i>Самостійна робота:</i> Побудувати алгоритм еволюційної стратегії / Theme 8. Distributed Transactions. <i>Individual work:</i> Build an algorithm for evolutionary strategy	2	2	7
9	Тема 9. Послідовність дій та консенсус <i>Самостійна робота:</i> Проаналізувати алгоритми консенсусу / Theme 9. Consistency and Consensus. <i>Individual work:</i> Analyze main algorithms of consensus	2	2	7
<i>Контрольна робота 2/ Control work 2</i>		1		
Частина III. Робота з даними/ Part III. Derived Data				
10	Тема 10. Пакетна обробка даних <i>Самостійна робота:</i> Побудувати застосування, що пакетно оброблює дані. / Theme 10. Batch Processing. <i>Individual work:</i> Develop application that work with data in batch processing paradigm	4	2	6
11	Тема 11. Поточкова обробка даних <i>Самостійна робота:</i> Порівняти пакетну і потокову обробку даних та визначити в яких ситуаціях який підхід найкраще підходить / Theme 11. Stream Processing.	1	2	5

<i>Individual work:</i> Analyze batch and stream data processing and provide summary for best cases			
<i>Контрольна робота 3</i>	1		
ВСЬОГО	28	10	80

Загальний обсяг 120 год., в тому числі/ Total duration 120 hours, namely:

Лекції/ Lectures – **28 год./h.**

Лабораторні роботи / Laboratory works – **10 год./h.**

Консультації/ Consultations - **2 год./h.**

Самостійна робота/ Individual work - **80 год. /h.**

9. Рекомендовані джерела

Основні / Main:

1. Э. Таненбаум, М. ван Стеен Распределённые системы, принципы и парадигмы. Питер, 2003
2. Martin Kleppmann Designing Data-Intensive Applications, OReilly 2017

Додаткові / Additional:

3. Д. Бэкон, Т. Харрис Операционные системы, параллельные и распределённые системы. Питер, 2004
4. В.А. Шехонцов Операционні системи. Київ, Видавнича група BHV, 2005