

**НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ  
ІНСТИТУТ ПРОБЛЕМ РЕЄСТРАЦІЇ ІНФОРМАЦІЇ**

ЗАТВЕРДЖУЮ



Директор ІПРІ НАН України  
академік НАН України

В.В. Петров

« 16 » 11 2021 р.

**ОСНОВИ ТЕОРІЇ ІНФОРМАЦІЙНОГО ПОШУКУ**

(назва навчальної дисципліни)

**РОБОЧА ПРОГРАМА  
кредитного модуля**

**ГАЛУЗЬ ЗНАНЬ** 12 «Інформаційні технології»  
**СПЕЦІАЛЬНІСТЬ** 122 «Комп'ютерні науки»  
**СПЕЦІАЛІЗАЦІЯ** «Інформаційні технології»

Ухвалено Вченою радою ІПРІ НАН України  
(протокол від « 16 » 11 2021 р. № 11)

Київ  
ІПРІ НАН України  
2021

Робоча програма кредитного модуля Основи теорії інформаційного пошуку  
складена відповідно до програми навчальної дисципліни Основи теорії  
інформаційного пошуку.

Розробник робочої програми:

Завідувач відділу, професор., д.т.н. Ланде Дмитро Володимирович  
(посада, науковий ступінь, вчене звання, прізвище, ім'я, по батькові)



(підпис)

## 1. Опис кредитного модуля

Рівень ВО, спеціальність, освітня програма, форма навчання	Загальні показники	Характеристика кредитного модуля
Рівень ВО третій (доктор філософії)	Основи теорії інформаційного пошуку	Лекції <u>16</u> год.
Спеціальність <i>122 Комп'ютерні науки</i>	Цикл професійної підготовки	Практичні (семінарські) <u>14</u> год.
Освітньо-наукової програми спеціальності 122 Комп'ютерні науки спеціалізації Інформаційні технології	Статус кредитного модуля обов'язковий	Лабораторні роботи -
		Самостійна робота <u>30</u> год.,
	Семестр <u>4</u>	Індивідуальне завдання <i>Не передбачено</i>
Форма навчання: денна	Кількість кредитів (годин): 2 (60)	Вид та форма семестрового контролю: <i>залік</i>

Предметом вивчення навчальної дисципліни «Основи теорії інформаційного пошуку» є питання, що відносяться до інформаційної структури Web-простору, повнотекстовим інформаційно-пошуковим системам, їх алгоритмічному та лінгвістичному забезпеченню, можливостям ранжирування, аналітичного узагальнення результатів пошуку, загальним закономірностям сучасного інформаційного простору.

Дисципліна складається з двох змістовних модулів:

1. Теоретико-методологічні основи інформаційного пошуку
2. Застосування методів інформаційного пошуку.

Вивчення дисципліни спирається на знання, отримані за програмою попередніх років навчання за спеціальністю 122 Комп'ютерні науки. Аспіранти мають досвід у таких областях математики, як теорія множин, вища алгебра, диференціальні рівняння та теорія ймовірностей, а також здатні оволодіти необхідними навичками для проведення практичних занять: мовами програмування Perl, Python програмуванням у веб-середовищі, прикладними програмами статистичної обробки даних.

## 2. Мета та завдання кредитного модуля

Метою викладання дисципліни є формування у слухачів наукового світогляду, цілісного уявлення про методологію наукового дослідження та навичок практичного застосування конкретних методів інформаційного пошуку у професійній діяльності, формування системи теоретичних і практичних знань в галузі інформаційних технологій, зокрема, при роботі з інформаційно-пошуковими

системами, серед яких сьогодні важлива роль відводиться пошуковими системами в мережі Інтернет.

2.1. Відповідно до вимог Освітньо-наукової програми третього (доктор філософії) рівня вищої освіти спеціальності 122 Комп'ютерні науки аспіранти після засвоєння навчальної дисципліни мають продемонструвати такі результати навчання:

**Здатності:**

- працювати з основними мережевими інформаційно-пошуковими системами;
- застосовувати сучасні інформаційні технології у різних видах професійної діяльності (ІК-2);
- знаходити, обробляти й аналізувати необхідну інформацію для рішення проблем й прийняття рішень (ІК-3);
- проводити теоретичні й експериментальні дослідження, математичне й комп'ютерне моделювання в галузі знань "Інформаційні технології" (ПК-3);
- застосовувати високопродуктивні технології розподілених систем та паралельних обчислень при вирішенні науково-практичних задач (ПК-5с);
- проектування та програмної реалізації методів інформаційного пошуку в інформаційних середовищах різноманітного призначення, систем управління бізнес-процесами, мереж Інтернету речей, сервіс-орієнтованих середовищ та систем високопродуктивних кластерних обчислень;
- вирішувати масштабні обчислювальні задачі у розподілених інтелектуальних середовищах та контролювати хід обчислень за допомогою спеціалізованого програмного забезпечення;
- вибирати адекватні методи інформаційного пошуку для вирішення конкретних задач прогнозування, керування, класифікації та інтелектуального аналізу даних.

2.2. Основні завдання кредитного модуля.

Згідно з вимогами програми навчальної дисципліни аспіранти після засвоєння кредитного модуля мають продемонструвати такі результати навчання:

**знання :**

- концептуальних засад і технологій роботи з великими даними;
- сучасних інформаційно-комунікаційних технологій, програмно-апаратних засобів проведення наукових досліджень;
- можливостей використання інформаційно-пошукових технологій для автоматизації експерименту, обробки даних, оформлення результатів досліджень;
- технологій інтелектуальних обчислень та аналізу даних;
- комп'ютерного моделювання та обробки даних, паралельних обчислень з використанням спеціалізованих програмних засобів;
- методів розподіленого моделювання складних об'єктів і систем в обчислювальному середовищі, застосувань технологій штучного інтелекту в розподілених обчисленнях, базових алгоритмів інтелектуального аналізу детермінованих та недетермінованих даних;
- з проектування математичного, лінгвістичного, інформаційного і програмного забезпечення інформаційних систем, з розроблення інформаційних систем, комплексів та мереж.

**уміння:**

- застосовувати знаннєво-орієнтовані мережеві інформаційні системи при вирішенні наукових та прикладних задач, пов'язаних з аналізом, моделюванням, прогнозуванням та управлінням економічних та соціальних процесів суспільства;
- використовувати інформаційно-комунікаційні технології, універсальні та спеціалізовані програмно-апаратні засоби;
- здійснювати автоматизацію експерименту, статистичну обробку даних, оформлення результатів досліджень засобами інформаційних технологій;
- застосовувати методи розподіленого моделювання складних об'єктів і систем, технології штучного інтелекту в розподілених обчисленнях, інтелектуальні обчислення, проектувати та програмно реалізовувати методи комп'ютерної обробки надвеликих за обсягом даних;
- проектувати математичне, лінгвістичне, інформаційне і програмне забезпечення інформаційних систем, розробляти інформаційні системи, комплекси та мережі
- використовувати сучасні інформаційно-пошукових технологій та відповідні інфраструктури програмних рішень;
- використовувати методи інформаційного поршуку для вирішення практичних задач.

**3. Структура кредитного модуля**

Назви розділів і тем	Кількість годин				
	Всього	у тому числі			
		Лекції	Практичні	Лабораторні	СРС
1	2	3	4	5	6
<b><i>Розділ 1. Теоретико-методологічні основи інформаційного пошуку</i></b>					
Тема 1. Сучасний інформаційний простір – об'єкт та продукт застосування комп'ютерних інформаційних технологій	7	2	1	-	4
Тема 2. Моделі інформаційного пошуку	8	2	2	-	4
Тема 3. Методи класифікації інформації	8	2	2	-	4
Разом за розділом 1	23	6	5	-	12
<b><i>Розділ 2. Застосування методів інформаційного пошуку</i></b>					
Тема 1. Інформаційно-пошукові системи.	6	2	1	-	3
Тема 2. Елементи кластерного аналізу.	8	2	2	-	4
Тема 3. Нейронні мережі	8	2	2	-	4
Тема 4. Моделі інформаційних потоків	8	2	2	-	4
Тема 5. Основи прикладної теорія складних мереж	7	2	2	-	3
Разом за розділом 2	37	10	9	-	18
<b>Всього годин</b>	<b>60</b>	<b>16</b>	<b>14</b>	<b>-</b>	<b>30</b>

#### 4. Лекційні заняття

№ з/п	Назва теми лекції та перелік основних питань
1	<b>Мережа Інтернет – історія та основні протоколи.</b> Визначення. Історія. Основні протоколи. Стандартизація. RFC. Чинники саморозвитку. Статистика мережі Інтернет. Визначення гіпертексту. Історія гіпертексту. Реалізації гіпертекстових технологій. Інтернет та WWW. Основні складові веб-технологій – HTML, HTTP, веб-браузери. Модель веб-простору. Статистична та динамічна складові веб-простору. Проблеми сучасного веб-простору. Концепція семантичного веб.
2	<b>Моделі інформаційного пошуку.</b> Булева модель. Векторно-просторова модель. Ймовірнісна модель. Гібридні моделі. Показники якості інформаційного пошуку. Метрики TREC.
3	<b>Інформаційно-пошукові системи.</b> Визначення. Історія технологій. Інформаційно-пошукові мови. Схема реалізації пошуку в інформаційно-пошукових системах з інвертованим індексом. Архітектура сучасних мережевих інформаційно-пошукових систем.
4	<b>Методи класифікації інформації.</b> Визначення. Задача розпізнавання. Машинне навчання. Метод k-найближчих сусідів. Лінійний класифікатор. ДНФ-метод. Метод опорних векторів (SVM). Оцінка якості класифікації. Застосування.
5	<b>Елементи кластерного аналізу.</b> Визначення кластерного аналізу. Детально методи латентного семантичного індексування (LSA). Метод k-means. Гібридні методи. Застосування.
6	<b>Нейронні мережі.</b> Структура штучного нейрону. Нейрон як комп'ютер. Перцептрон. Нейронні мережі. Існуючі моделі нейронних мереж. Нейронна мережі, що вирішує завдання класифікації повідомлень. Машинне навчання. Глибоке навчання.
7	<b>Моделі інформаційних потоків.</b> Поняття інформаційних потоків. Лінійна, експонентна, логістична моделі. Кореляція між елементами інформаційних потоків. Моделювання інформаційних потоків як клітинних автоматів.
8	<b>Основи прикладної теорії складних мереж.</b> Основні показники мереж та вузлів. Близькі та далекі зв'язки. Концепція малих світів. Перколяція. Структура веб-простору. Ранжирування результатів пошуку з урахуванням структури мережі. Алгоритми HITS і PageRank.

#### 5. Семінарські заняття

Основні завдання циклу практичних занять полягають у набутті аспірантами практичних навичок з використання програмного забезпечення інформаційно-пошукових систем.

№ з/п	Назва теми заняття
1	Протоколи мережі Інтернет. Технологічне середовище проведення практичних занять.
2	Модель інформаційного веб-простору. Робота з пакетом статистичної

	обробки даних.
3	Моделі інформаційного пошуку. Булева модель.
4	Інформаційно-пошукові системи. Побудова текстових наборів даних інформаційно-пошукової системи.
5	Інформаційно-пошукові системи. Розробка алгоритму пошуку з використанням логічних операторів в інформаційно-пошуковій системі з інвертованим індексом.
6	Методи класифікації інформації. Метод опорних векторів.
7	Розробка правил моделювання інформаційних потоків за допомогою клітинних автоматів.

### 5. Практичні заняття

Практичних занять не передбачено.

### 6. Лабораторні заняття

Лабораторних занять не передбачено.

### 7. Самостійна робота

№ з/п	Назви тем і питань, що виносяться на самостійне опрацювання та посилання на навчальну літературу	Кількість годин СРС
1	<b>Сучасний інформаційний простір – об'єкт та продукт застосування комп'ютерних інформаційних технологій.</b> Стандартизація, чинники саморозвитку, статистика мережі Інтернет. Реалізації гіпертекстових технологій. Інтернет та WWW. Основні складові веб-технологій. Модель веб-простору. Концепція семантичного веб.	4
2	<b>Моделі інформаційного пошуку.</b> Булева модель. Векторно-просторова модель. Ймовірнісна модель. Гібридні моделі. Показники якості інформаційного пошуку. Метрики TREC: Повнота, точність F-міра.	4
3	<b>Методи класифікації інформації.</b> Метод k-найближчих сусідів. Лінійний класифікатор. ДНФ-метод. Метод опорних векторів (SVM). Оцінка якості класифікації.	4
4	<b>Інформаційно-пошукові системи.</b> Схема реалізації пошуку в інформаційно-пошукових системах з інвертованим індексом. Архітектура сучасних мережевих інформаційно-пошукових систем.	3
5	<b>Елементи кластерного аналізу.</b> Метод латентного семантичного індексування (LSA). Метод k-means. Метод ієрархічного агломеративного групування.	4
6	<b>Нейронні мережі.</b> Перцептрон. Нейронні мережі. Існуючі моделі нейронних мереж. Нейронна мережі, що вирішує завдання класифікації повідомлень. Машинне навчання. Глибоке навчання.	4

7	<b>Моделі інформаційних потоків.</b> Лінійна, експонентна, логістична моделі. Кореляція між елементами інформаційних потоків. Моделювання інформаційних потоків як клітинних автоматів – модель дифузії інформації.	4
8	<b>Основи прикладної теорія складних мереж.</b> Концепція малих світів.. Структура веб-простору. Ранжирування результатів пошуку з урахуванням структури мережі. Алгоритми HITS і PageRank.	3

### 8. Індивідуальні завдання

Індивідуальних завдань не передбачено.

### 9. Контрольні роботи

Передбачається одна модульна контрольна робота, метою якої є перевірка та закріплення набутих аспірантами знань. Варіант контрольної роботи містить два теоретичні питання.

### 10. Рейтингова система оцінювання результатів навчання

Оцінка з дисципліни виставляється за багатобальною системою, з подальшим перерахуванням у 4-бальну.

2. Максимальна кількість балів з дисципліни дорівнює 100.

3. Нарахування балів по окремих видах робіт:

Рейтинг аспіранта з кредитного модуля складається з балів, що він отримав за:

- 1) виконання практичних робіт;
- 2) написання контрольної роботи (МКР);

#### Система рейтингових (вагових) балів та критерії оцінювання

#### 1. Виконання практичних робіт

Оцінюються 8 робіт, передбачених робочою програмою. Максимальний ваговий бал глр =64

Сума вагових балів практичних робіт:

№ п.р.	Назва практичної роботи	Максимальний ваговий бал
1	Протоколи мережі Інтернет. Технологічне середовище проведення практичних занять.	9
2	Модель інформаційного веб-простору. Робота з пакетом статистичної обробки даних.	9
3	Моделі інформаційного пошуку. Булева модель.	9
4	Інформаційно-пошукові системи. Побудова текстових наборів даних інформаційно-пошукової системи.	9
5	Інформаційно-пошукові системи. Розробка алгоритму пошуку з використанням логічних операторів в інформаційно-пошуковій системі з інвертованим індексом.	9
6	Методи класифікації інформації. Метод опорних векторів.	9



7	Розробка правил моделювання інформаційних потоків за допомогою клітинних автоматів.	9
Разом		63

Оцінювання практичних робіт:

–якщо робота виконана невчасно знімається 10-30% від максимальної кількості балів (кількість процентів залежить від терміну запізнення);

–якщо робота виконана не самостійно та простежується не індивідуальне виконання то знімається 50% від максимальної кількості балів;

–якщо в програмі не витримані основні правила створення програмних продуктів (модульність, дружній інтерфейс, наявність коментарів та т.п.) знімається 5%.

## 2. Модульний контроль

На одному з лекційних занять проводиться модульна контрольна робота: Максимальний ваговий бал  $r_{МКР} = 11$ .

Оцінювання модульної контрольної роботи виконується наступним чином:

–якщо на всі питання дані повні та чітко аргументовані відповіді, контрольна виконана охайно, з дотримання основних правил, то виставляється 9 - 11 балів;

–якщо методика виконання запропонованого завдання розроблена вірно, але допущені неprincipові помилки у теоретичному описі або розрахунках, то виставляється 6 - 8 балів;

–від 3 до 5 балів нараховується, якщо методика виконання завдання розроблена в основному вірно, але допущені деякі з наступних помилок: помилки у представленні вихідних даних, не обґрунтовані теоретичні рішення, помилки у методиці розрахунків;

–нижче 3 балів нараховується, якщо завдання не виконане або допущені грубі помилки.

## 3. Залік

Залік відбувається у письмовій формі. Максимальна оцінка за залік складає  $r_{ЕК} = 25$  балів.

### *Розрахунок шкали (R) рейтингу:*

Сума вагових балів контрольних заходів протягом семестру складає:

$$R = 64 + 11 + 25 = 100 \text{ балів}$$

Таким чином, рейтингова шкала з кредитного модуля складає 100 балів.

Умови допуску до заліку: зарахування всіх лабораторних робіт, а також стартовий рейтинг  $r \geq 40$  балів.

Для отримання аспірантом відповідних оцінок (ECTS та традиційних) його рейтингова оцінка **R** переводиться згідно таблиці:

Шкала оцінювання:

За 100 – бальною шкалою	За національною шкалою
90 – 100	Зараховано
85 – 89	
75 – 84	
65 – 74	

60 – 64	
1 – 59	не зараховано

1

## 11. Методичні рекомендації

Для кращого засвоєння матеріалу дисципліни рекомендується використовувати на лекціях мультимедійні засоби навчання, які дозволяють інтенсифікувати навчальний процес, стимулювати розвиток мислення та уяви аспірантів, збільшувати обсяг навчального матеріалу для творчого засвоєння і використання його аспірантами, викликати зацікавленість та позитивне ставлення до навчання.

Методика побудована таким чином, що матеріал майже кожної лекції закріплюється виконанням завдання комп'ютерного практикуму. Завдання аспіранти отримують заздалегідь і на аудиторному занятті під керівництвом викладача виправляють помилки в разі їх наявності та відповідають на запитання щодо програмної реалізації та теоретичних засад роботи.

Якість самостійної роботи перевіряється на заняттях комп'ютерного практикуму.

## 12. Рекомендована література

### 12.1. Базова:

1. Ландэ Д.В., Снарский А.А., Безсуднов И.В. Интернетика: Навигация в сложных сетях: модели и алгоритмы - М.: Либликом (Editorial URSS), 2009. - 264 с.
2. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа. - М.: Диалектика, 2005. - 272 с
3. Додонов А.Г., Ландэ Д.В., Путятин В.Г. Компьютерные сети и аналитические исследования. - К.: ИПРИ НАН Украины, 2014. - 486 с.
4. Арнольд В.И. Аналитика и прогнозирование: математический аспект. // Научно-техническая информация. Сер. 1. Вып. 3. - 2003. - С. 1-10.
5. Ландэ Д.В. Основы интеграции информационных потоков - К.: Инжиниринг, 2006. - 240 с. (<http://dwl.kiev.ua/art/monogr-osnov/spusk3.pdf>)
6. Ландэ Д.В. Фрактальные свойства тематических информационных потоков из Интернет // Реєстрація, зберігання і обробка даних, 2006, Т. 8, № 2.– С. 93 - 99.
7. Мандельброт Б. Фрактальная геометрия природы. – М.: Институт компьютерных исследований, 2002 г. - 656 с.
8. Фон Нейман Дж. Теория самовоспроизводящихся автоматов - М.: Мир, 1971.
9. Тоффоли Т., Марголус Н. Машины клеточных автоматов. М.: Мир, 1991.
10. Федер Е. Фракталы -М.: Мир, 1991. - 254 с.
11. Ricardo Baeza-Yates and Berthier Ribeiro-Neto. Modern Information Retrieval. ACM Press, 1999. – 513 p.
12. S. Brin and L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. WWW7, - 1998.
13. Tim Berners-Lee, James Hendler, Ora Lassila. The Semantic Web. Scientific American,

2001 (<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>)

14. Michael W. Berry. Survey of Text Mining. Clustering, Classification, and Retrieval. - Springer-Verlag, 2004. - 244 p.
15. J.M. Kleinberg. Authoritative sources in a hyperlink environment. // In Processing of ACM-SIAM Symposium on Discrete Algorithms, 1998, 46(5):604–632.
16. Newman M.E.J. The structure and function of complex networks. // SIAM Review. - 2003. - Vol. 45. pp. 167–256.

### **12.2. Допоміжна:**

17. Арнольд В.И. Обыкновенные дифференциальные уравнения. М.: Наука, 1971, - 240 с.
18. Иванов С.А. Стохастические фракталы в Информатике // Научно-техническая информация. Сер. 2, 2002. - № 8. - С. 7-18.
19. S. Wolfram. A New Kind of Science. Champaign, IL: Wolfram Media Inc., 2002. – 1197 pp.
20. S. Wolfram ed. Theory and Applications of Cellular Automats. Singapore: World Scientific. 1986.
21. Chakrabarti Soumen. Mining the web. Discovery knowledge from hypertext data. - Publisher: Morgan Kaufmann, 2002. - 344 p.
22. The Deep Web: Surfacing Hidden Value, 2000 BrightPlanet.com LLC, 35 p. (<http://www.dad.be/library/pdf/BrightPlanet.pdf>)
23. Watts D.J., Strogatz S.H. Collective dynamics of "small-world" networks. // Nature. - 1998. - Vol. 393. pp. 440–442.