



# ІНФОРМАЦІЙНО-ПОШУКОВІ СИСТЕМИ ТА СЕРВІСИ

## Робоча програма навчальної дисципліни (Силабус)

### Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>121 Інженерія програмного забезпечення</i>
Освітня програма	<i>Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем</i>
Статус дисципліни	<i>Нормативна</i>
Форма навчання	<i>Очна (денна)</i>
Рік підготовки, семестр	<i>1 рік підготовки, 1 семестр</i>
Обсяг дисципліни	<i>Лекції: 36 год., комп'ютерний практикум: 36 год., самостійна робота: 78 год.</i>
Семестровий контроль/ контрольні заходи	<i>Екзамен, модульна контрольна робота, календарний контроль</i>
Розклад занять	<i>Згідно розкладу на осінній семестр поточного навчального року (<a href="http://roz.kpi.ua/">http://roz.kpi.ua/</a>)</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: к.т.н., асистент, Погорелов В.В., <a href="mailto:volodymyr.pogorelov@gmail.com">volodymyr.pogorelov@gmail.com</a> Комп'ютерний практикум: аспірант, асистент, Северін А.І., <a href="mailto:severinandrey97@gmail.com">severinandrey97@gmail.com</a></i>
Розміщення курсу	<i>Google classroom: <a href="https://classroom.google.com/u/0/c/MzE5ODIyMjA4MTA3?hl=ua">https://classroom.google.com/u/0/c/MzE5ODIyMjA4MTA3?hl=ua</a></i>

### Програма навчальної дисципліни

#### 1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Вивчення дисципліни «Інформаційно-пошукові системи та сервіси» дозволяє сформувати у здобувачів освіти компетенції, необхідні для розв'язання практичних задач професійної діяльності, пов'язаної із розробленням програмного забезпечення для інформаційно-пошукових систем.

**Метою** вивчення дисципліни «Інформаційно-пошукові системи та сервіси» є формування у здобувачів освіти здатностей самостійно проектувати та розроблювати програмне забезпечення, яке реалізує методи та алгоритми пошуку даних в інформаційно-пошукових системах та сервісах.

**Предметом** дисципліни «Інформаційно-пошукові системи та сервіси» є методи, алгоритми та моделі, що використовуються для розроблення інформаційно-пошукових систем.

Вивчення дисципліни «Інформаційно-пошукові системи та сервіси» сприяє формуванню у здобувачів освіти **фахових компетентностей (ФК)**, необхідних для розв'язання практичних

задач професійної діяльності, пов'язаної з розробленням, вдосконаленням та експлуатацією інформаційно-пошукових систем:

**ФК03** Здатність проектувати архітектуру програмного забезпечення, моделювати процеси функціонування окремих підсистем і модулів.

**ФК05** Здатність розробляти, аналізувати та застосовувати специфікації, стандарти, правила і рекомендації в сфері інженерії програмного забезпечення.

**ФК10** Здатність проектувати складні мультимедійні та інформаційно-пошукові системи.

**ФК11** Здатність проектувати та конструювати, впроваджувати та підтримувати веборієнтовані програмні системи для реалізації нових методів пошуку інформації.

**ФК13** Здатність впроваджувати та підтримувати інформаційні системи.

Вивчення дисципліни «Інформаційно-пошукові системи та сервіси» сприяє формуванню у студентів наступних **програмних результатів навчання (ПРН)** за освітньою програмою:

**ПРН04** Виявляти інформаційні потреби і класифікувати дані для проектування програмного забезпечення.

**ПРН18** Знати теоретичні засади, що лежать в основі методів досліджень інформаційних систем та програмного забезпечення, методології проведення досліджень та обчислювальних експериментів.

**ПРН20** Знати принципи побудови програмних інформаційно-пошукових систем.

**ПРН22** Вміти модифікувати існуючі та розробляти нові методи і алгоритми пошуку мультимедійних даних в інформаційно-пошукових системах, з урахуванням особливостей предметної області.

**ПРН23** Знати та вміти застосовувати на практиці спеціалізовані шаблони проектування інформаційно-пошукових систем.

**ПРН24** Вміти проектувати та розробляти мультиагентні інформаційно-пошукові системи.

**ПРН25** Вміти проектувати та розробляти розподілені та централізовані інформаційно-пошукові системи.

**ПРН26** Знати та вміти використовувати засоби інформаційного ущільнення алфавітно-цифрових даних.

## **2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)**

Успішному вивченню дисципліни «Інформаційно-пошукові системи та сервіси» передують вивчення дисциплін «Математичний аналіз», «Лінійна алгебра та аналітична геометрія», «Теорія імовірностей», «Структури даних та алгоритми», «Програмування» та «Програмне забезпечення інформаційно-пошукових систем» навчального плану підготовки бакалаврів за спеціальністю 121 Інженерія програмного забезпечення.

Отримані при засвоєнні дисципліни «Інформаційно-пошукові системи та сервіси» теоретичні знання та практичні уміння забезпечують успішне виконання курсових проєктів та магістерських дисертацій за спеціальністю 121 Інженерія програмного забезпечення.

## **3. Зміст навчальної дисципліни**

Дисципліна «Інформаційно-пошукові системи та сервіси» передбачає вивчення таких тем:

Тема 1. Вступ в пошукові системи та сервіси

Тема 2. Інвертовані індекси

Тема 3. Веб-граф та аналіз посилань

Тема 4. Інфраструктура за межами індексу

Тема 5. Користувачі та реклама

Модульна контрольна робота

#### 4. Навчальні матеріали та ресурси

##### **Базова література:**

1. Алгоритми пошуку в інформаційних системах: метод. реком. / О.Л. Сухий, В.М. Міленін, В.М. Тарадайнік. – К.: Інститут обдарованої дитини НАПН України, 2015. – 70 с.
2. Руденко В.Д. Бази даних в інформаційних системах К.: Фенікс, 2010,- 235 с.
3. Нестеренко О.В. Інтелектуальні системи підтримки прийняття рішень: навч. посібн./ О.В. Нестеренко, О.І. Савенков, О.О. Фаловський. За ред. П.І. Бідюка. – Київ: Національна академія управління. – 2016. – 188 с.
4. Вступ до інженерії програмного забезпечення: навч. посібник / Є.В. Левус, Н.Б. Мельник – Львів: Видавництво Львівської політехніки, 2017. – 280 с.

##### **Додаткова література:**

5. S. Abiteboul, M. Preda, and G. Cobena. Adaptive on-line page importance computation. In Proc. 12th International WWW Conference (WWW2003), Budapest, Hungary, pages 280–290, 2003.
6. G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6):734–749, 2005.
7. G. Aggarwal, S. M. Muthukrishnan, D. Pal, and M. Pal. General auction mechanisms for search advertising. In Proc. 18th International World Wide Web Conference (WWW'2009), pages 241–250, April 2009.
8. S. Amer-Yahia, M. D. Choudhury, M. Feldman, N. Golbandi, R. Lempel, and C. Yu. Automatic construction of travel itineraries using social breadcrumbs. In Proc. 21st ACM Conference on Hypertext and Hypermedia (Hypertext'2010), pages 35–44, June 2010.
9. C. Anderson. The Long Tail - Why the Future of Business is Selling Less of More. Hyperion Books, New York NY, 2006.
10. V. Anh and A. Moffat. Index compression using fixed binary codewords. In Proc. of the 15th Int. Australasian Database Conference, pages 61–67, 2004. 4
11. A. Arasu, J. Cho, H. Garcia-Molina, A. Paepcke, and S. Raghavan. Searching the web. ACM Transactions on Internet Technology, 1(1):2–43, 2001.
12. R. Baeza-Yates. Graphs from search engine queries. In Proc. 33rd conference on Current Trends in Theory and Practice of Computer Science, SOFSEM'07, pages 1–8, 2007.
13. R. Baeza-Yates, A. Gionis, F. Junqueira, V. Murdock, V. Plachouras, and F. Silvestri. The Impact of Caching on Search Engines. In SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, 2007. ACM Press.
14. R. Baeza-Yates, C. Hurtado, and M. Mendoza. Improving search engines by query clustering. J. Am. Soc. Inf. Sci. Technol., 58:1793–1804, October 2007.
15. R. Baeza-Yates, F. Junqueira, V. Plachouras, and H. F. Witschel. Admission Policies for Caches of Search Engine Results. In SPIRE, 2007.

Використати для опанування практичних умінь дисципліни. Матеріали знаходяться у вільному доступі в Інтернеті.

**Навчальний контент**

**5. Методика опанування навчальної дисципліни (освітнього компонента)**

№ з/п	Тип навчального заняття	Опис навчального заняття
<i>Тема 1. Вступ в пошукові системи та сервіси</i>		
1	<i>Лекція 1. Зміст курсу, вступ до пошукових систем, технічний огляд компонентів пошукової системи</i>	<i>Огляд змісту курсу, вступ до пошукових систем. Вступ до інформаційного пошуку: Булева модель, векторна модель, показник TF/IDF. Технічний огляд компонентів пошукової системи  Завдання на СРС: п.6 №1.</i>
2	<i>Лекція 2. Імовірнісний пошук інформації, лема Неймана-Пірсона</i>	<i>Імовірнісний пошук інформації. Лема Неймана-Пірсона. Мовні моделі в інформаційному пошуку  Завдання на СРС: п.6 №2.</i>
<i>Тема 2. Інвертовані індекси</i>		
3	<i>Лекція 3. Основи: поняття інвертованого індексу; ефективна побудова індексу; операції, що він підтримує; додаткове корисне навантаження, що зазвичай зберігається в пошукових системах; супровідний лексикон</i>	<i>Основи: поняття інвертованого індексу. Методи ефективної побудови індексу. Операції, що підтримує індекс. Додаткове корисне навантаження, що зазвичай зберігається в пошукових системах. Супровідний лексикон B-Tree, Min-Heap  Завдання на СРС: п.6 №3.</i>
4	<i>Лекція 4. Схеми оцінки запитів: термін за терміном (term-at-a-time) і документ за документом (doc-at-a-time), купи результатів, раннє завершення/спрощення, WAND</i>	<i>Схеми оцінки запитів: термін за терміном (term-at-a-time) і документ за документом (doc-at-a-time). Ідентифікація near-дубліката сторінки. Купи результатів, раннє завершення/спрощення, WAND  Завдання на СРС: п.6 №4.</i>
5	<i>Лекція 5. Стиснення індексу та зміна порядку документів</i>	<i>Стиснення індексу та зміна порядку документів. Пошукова бібліотека Apache Lucene (необхідна умова для домашнього завдання)  Завдання на СРС: п.6 №5.</i>
6	<i>Лекція 6. Архітектури розподілених індексів: глобальні/локальні схеми, комбінаторні проблеми, пов'язані з розподілом даних, кластерна архітектура Google</i>	<i>Архітектури розподілених індексів. Глобальні/локальні схеми, комбінаторні проблеми, пов'язані з розподілом даних, кластерна архітектура Google.  Завдання на СРС: п.6 №6.</i>
7	<i>Комп'ютерний практикум 1. Побудова інвертованих індексів</i>	<i>Завдання: Проаналізувати програмні засоби та побудувати індекс в системі інформаційного пошуку.  Завдання на СРС: п.6 №7.</i>

*Тема 3. Веб-граф та аналіз посилань*

8	<i>Лекція 7. Структура веб-графа: степеневі закони, структура Vow-tie, самоподібність</i>	<i>Структура веб-графа: степеневі закони, структура Vow-tie, самоподібність Завдання на СРС: п.6 №8.</i>
9	<i>Лекція 8. Основи аналізу посилань: PageRank Google, Kleinberg HITS, короткий огляд теорії Перрона-Фробеніуса та ергодичності</i>	<i>Основи аналізу посилань: PageRank Google (чутливий до теми), Kleinberg HITS, короткий огляд теорії Перрона-Фробеніуса та ергодичності Завдання на СРС: п.6 №9.</i>
10	<i>Комп'ютерний практикум 2. Аналіз посилань</i>	<i>Завдання: За допомогою програмних засобів реалізувати модуль для аналізу посилань для інформаційної системи. Завдання на СРС: п.6 №10.</i>
11	<i>Лекція 9. Стабільність та подібність схем на основі зв'язків, ТКС ефект</i>	<i>Стабільність та подібність схем на основі зв'язків, ТКС ефект. Еволюційні моделі веб-графа Завдання на СРС: п.6 №11.</i>

*Тема 4. Інфраструктура за межами індексу*

12	<i>Лекція 10. Краулери. Призначення та архітектура, оптимізація порядку сканування, обчислення метрик важливості під час сканування</i>	<i>Краулери. Призначення та архітектура, оптимізація порядку сканування, обчислення метрик важливості під час сканування. Фільтри Блума Завдання на СРС: п.6 №12.</i>
13	<i>Лекція 11. Ефективне кешування та попередня вибірка результатів запиту</i>	<i>Ефективне кешування та попередня вибірка результатів запиту Завдання на СРС: п.6 №13.</i>

*Тема 5. Користувачі та реклама*

14	<i>Лекція 12. Комп'ютерна реклама: моделі та визначення. СРМ, СРС, СРА; спонсорований пошук (adwords), відповідність вмісту (adsense), медійна реклама</i>	<i>Комп'ютерна реклама: моделі та визначення. СРМ, СРС, СРА. Спонсорований пошук (adwords), відповідність вмісту (adsense), медійна реклама Завдання на СРС: п.6 №14.</i>
15	<i>Лекція 13. Комп'ютерна реклама: механізми аукціону</i>	<i>Комп'ютерна реклама (ТВД): моделі та визначення. СРМ, СРС, СРА; спонсорований пошук (adwords), відповідність вмісту (adsense), медійна реклама Завдання на СРС: п.6 №15.</i>
16	<i>Лекція 14. Видобуток і перехоплення неявного</i>	<i>Видобуток і перехоплення неявного контенту, створеного користувачами. Аналіз журналу запитів</i>

	<i>контенту, створеного користувачами</i>	<i>Завдання на СРС: п.6 №16.</i>
<i>17</i>	<i>Лекція 15. Виконання завдань і допомога в пошуку – від виправлення орфографії та простих ярликів до мультимедійних матеріалів, мешапів, завершення запитів і аспектів</i>	<i>Виконання завдань і допомога в пошуку – від виправлення орфографії та простих ярликів до мультимедійних матеріалів, мешапів, завершення запитів і аспектів. Завдання на СРС: п.6 №17.</i>
<i>18</i>	<i>Лекція 15. Long Tail, системи рекомендацій і спільна фільтрація</i>	<i>Long Tail, системи рекомендацій і спільна фільтрація. Контекстно-залежний пошук і моделювання користувачів Завдання на СРС: п.6 №18.</i>
<i>19</i>	<i>Комп'ютерний практикум 3. Побудова системи контекстно-залежного пошуку</i>	<i>Завдання: побудувати систему контекстно-залежного пошуку. Завдання на СРС: п.6 №19.</i>
<i>Модульна контрольна робота</i>		

## **6. Самостійна робота студента/аспіранта**

*Дисципліна «Інформаційно-пошукові системи та сервіси» ґрунтується на самостійних підготовках до аудиторних занять на теоретичні та практичні теми.*

<i>№ з/п</i>	<i>Назва теми, що виноситься на самостійне опрацювання</i>	<i>Кількість годин</i>	<i>Література</i>
<i>1</i>	<i>Підготовка до лекції 1</i>	<i>1</i>	<i>1; 2; 3; 8</i>
<i>3</i>	<i>Підготовка до лекції 2</i>	<i>1</i>	<i>1; 4, 7</i>
<i>4</i>	<i>Підготовка до лекції 3</i>	<i>1</i>	<i>1; 3; 5; 9; 10</i>
<i>5</i>	<i>Підготовка до лекції 4</i>	<i>1</i>	<i>11-13</i>
<i>6</i>	<i>Підготовка до лекції 5</i>	<i>1</i>	<i>3; 6; 10</i>
<i>7</i>	<i>Підготовка до лекції 6</i>	<i>1</i>	<i>3; 8; 10</i>
<i>8</i>	<i>Підготовка до комп'ютерного практикуму 1</i>	<i>5</i>	<i>11-13</i>
<i>9</i>	<i>Підготовка до лекції 6</i>	<i>1</i>	<i>2; 5</i>
<i>10</i>	<i>Підготовка до лекції 7</i>	<i>1</i>	<i>2; 9</i>
<i>11</i>	<i>Підготовка до лекції 8</i>	<i>1</i>	<i>11-13</i>
<i>12</i>	<i>Підготовка до комп'ютерного практикуму 2</i>	<i>5</i>	<i>2; 6</i>
<i>13</i>	<i>Підготовка до лекції 9</i>	<i>1</i>	<i>1; 2; 3</i>
<i>14</i>	<i>Підготовка до лекції 10</i>	<i>1</i>	<i>11-13</i>
<i>15</i>	<i>Підготовка до лекції 11</i>	<i>1</i>	<i>4; 5</i>
<i>16</i>	<i>Підготовка до лекції 12</i>	<i>1</i>	<i>4; 5</i>

17	Підготовка до лекції 13	1	9-10
18	Підготовка до лекції 14	1	4, 12
19	Підготовка до лекції 15	1	4, 10
20	Підготовка до комп'ютерного практикуму 3	5	3, 11-13
27	Підготовка до модульної контрольної роботи	12	1-15
28	Підготовка до екзамену	36	1-15

## Політика та контроль

### 7. Політика навчальної дисципліни (освітнього компонента)

*Відвідування лекційних занять є обов'язковим.*

*Відвідування занять комп'ютерного практикуму може бути епізодичним та за потреби консультації/захисту робіт комп'ютерного практикуму.*

*Правила поведінки на заняттях: активність, повага до присутніх, відключення телефонів.*

*Дотримання політики академічної доброчесності.*

*Правила захисту робіт комп'ютерного практикуму: роботи повинні бути зроблені відповідно до поставлених задач та згідно з варіантом.*

### 8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

*Протягом семестру студенти виконують 3 комп'ютерних практикуми. **Максимальна кількість балів** за кожний комп'ютерний практикум: 10 балів.*

*Бали нараховуються за:*

- якість виконання комп'ютерного практикуму: 0-6 бали;
- відповідь під час захисту комп'ютерного практикуму: 0-2 бали;
- своєчасне представлення роботи до захисту: 0-2 бали.

*Критерії оцінювання якості виконання:*

- 6 балів – робота виконана якісно, в повному обсязі;
- 3-5 бали – робота виконана якісно, в повному обсязі, але має недоліки;
- 1-2 балів – робота виконана в повному обсязі, але містить суттєві помилки;
- 0 балів – робота виконана не в повному обсязі.

*Критерії оцінювання відповіді:*

- 2 бали – відповідь повна, добре аргументована;
- 1 бали – у відповіді є суттєві помилки;
- 0 балів – немає відповіді або відповідь невірна.

*Критерії оцінювання своєчасності представлення роботи до захисту:*

- 2 бали – робота представлена до захисту не пізніше вказаного терміну;
- 0 балів – робота представлена до захисту пізніше вказаного терміну.

**Максимальна кількість балів за виконання та захист комп'ютерних практикумів:**

*10 балів × 3 комп. практ. = 30 балів.*

*Завдання на **модульну контрольну роботу** складається з 3 теоретичних та 2 практичних запитань. Відповідь на кожне запитання оцінюється 4 балами.*

*Критерії оцінювання кожного запитання контрольної роботи:*

- 4 бали – відповідь вірна, повна, добре аргументована;
- 2-3 балів – в цілому відповідь вірна, але має недоліки;
- 1 бали – у відповіді є суттєві помилки;

0 балів – немає відповіді або відповідь невірна.

**Максимальна кількість балів за модульну контрольну роботу:**

4 бали × 5 запитань = 20 балів.

**Рейтингова шкала з дисципліни дорівнює:**

$R = R_c = R_{\text{ком.практ}} + R_{\text{МКР}} + R_{\text{екзамен}} = 30 \text{ балів} + 20 \text{ балів} + 50 \text{ балів} = 100 \text{ балів}.$

Календарний контроль: проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

На першій атестації (8-й тиждень) студент отримує «зараховано», якщо його поточний рейтинг не менше 10 балів (50 % від максимальної кількості балів, яку може отримати студент до першої атестації).

На другій атестації (14-й тиждень) студент отримує «зараховано», якщо його поточний рейтинг не менше 20 балів (50 % від максимальної кількості балів, яку може отримати студент до другої атестації).

Семестровий контроль: екзамен

Умови допуску до семестрового контролю:

При семестровому рейтингу ( $R_c$ ) не менше 30 балів та зарахуванні усіх робіт комп'ютерного практикуму, студент має допуск до екзамену. Після складання екзамену виставляється оцінка відповідно до таблиці (Таблиця відповідності рейтингових балів оцінкам за університетською шкалою).

Необхідною умовою допуску до екзамену є виконання та захист комп'ютерного практикуму.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Кількість балів	Оцінка
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

**9. Додаткова інформація з дисципліни (освітнього компонента)**

Перелік питань, які виносяться на семестровий контроль, наведено у Додатку 1.

**Робочу програму навчальної дисципліни (силабус):**

Складено к.т.н., асистент, Погорелов В.В.; аспірант, асистент, Северін А.І.

Ухвалено кафедрою ПЗКС (протокол № 13 від 22.06.22)

Погоджено Методичною комісією факультету прикладної математики (протокол № 9 від 24.06.22)



*Додаток 1. Перелік питань, які виносяться на семестровий контроль*

- 1. Імовірнісний пошук інформації, лема Неймана-Пірсона*
- 2. Поняття інвертованого індексу; ефективна побудова індексу.*
- 3. Додаткове корисне навантаження, що зазвичай зберігається в пошукових системах; супровідний лексикон.*
- 4. Схеми оцінки запитів: термін за терміном (term-at-a-time) і документ за документом (doc-at-a-time).*
- 5. Компоненти пошукової системи.*
- 6. Ефективне кешування та попередня вибірка результатів запиту.*
- 7. Кластерна архітектура Google.*
- 8. Архітектури розподілених індексів: глобальні/локальні схеми, комбінаторні проблеми, пов'язані з розподілом даних.*
- 9. Структура веб-графа: степеневі закони, структура Vow-tie, самоподібність.*
- 10. Основи аналізу посилань: PageRank Google, Kleinberg HITS, короткий огляд теорії Перрона-Фробеніуса та ергодичності.*
- 11. Стабільність та подібність схем на основі зв'язків, ТКС ефект.*
- 12. Краулери. Призначення та архітектура, оптимізація порядку сканування, обчислення метрик важливості під час сканування.*
- 13. Комп'ютерна реклама: моделі та визначення. CPM, CPC, CPA; спонсорований пошук (adwords), відповідність вмісту (adsense), медійна реклама.*
- 14. Видобуток і перехоплення неявного контенту, створеного користувачами.*
- 15. Стиснення індексу та зміна порядку документів.*
- 16. Long Tail, системи рекомендацій і спільна фільтрація.*
- 17. Комп'ютерна реклама: механізми аукціону.*
- 18. Призначення та архітектура, оптимізація порядку сканування, обчислення метрик важливості під час сканування.*
- 19. Видобуток і перехоплення неявного контенту, створеного користувачами.*
- 20. Пошукові задачі, засоби та технології інформаційного пошуку.*