

МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ

Харківський національний університет внутрішніх справ

Кафедра інформаційних технологій та кібербезпеки, факультет № 4

МЕТОДИЧНІ МАТЕРІАЛИ

ДО СЕМІНАРСЬКИХ ЗАНЯТЬ

з навчальної дисципліни «Моделі, методи та засоби аналітичної обробки

великих масивів даних»

вибіркових компонент

освітньої програми другого(магістерського) рівня вищої освіти

125 «Кібербезпека» («Безпека інформаційних та комунікаційних систем»)

м. Харків 2020

ЗАТВЕРДЖЕНО

Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від 23.09.2020 № 9

СХВАЛЕНО

Вченою радою факультету № 4
Протокол від 16.09.2020 № 5

ПОГОДЖЕНО

Секцією Науково-методичної ради
ХНУВС з технічних дисциплін

Протокол від 18.09.2020 № 5

Розглянуто на засіданні кафедри інформаційних технологій та кібербезпеки
(протокол від 15.09.2020 № 16)

Розробники:

1. Професор. кафедри, к.т.н., доцент Струков В.М.

Рецензенти:

1. д.т.н., професор Зацеркляний М.М.,
2. доцент кафедри програмної інженерії ХНУРЕ, кандидат технічних наук,
доцент Лановий О.Ф.

1. Розподіл часу навчальної дисципліни за темами

Номер та назва навчальної теми	Кількість годин, відведених на вивчення навчальної дисципліни					Вид контролю	
	Всього	з них:					
		лекції	Семінарські заняття	Практичні заняття	Лабораторні заняття		Самостійна робота
Семестр № 1							
Тема № 1. Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері Проблема обробки великих масивів даних.	18	4	4			10	к/р
Тема № 2: Задачі і етапи опрацювання великих даних.	12	2	2			8	
Тема № 3: Елементи теорії множин і теорії графів.	18	4	2			12	к/р
Тема № 4.Статистичне дослідження великих даних.	36	4		8		24	
Тема № 5: Алгоритми ієрархічної кластеризації.	30	4	2	4		20	
Тема № 6: Методи і алгоритми чіткої кластеризації.	34	4	2	4		24	
Тема № 7: Інструментальні засоби обробки Великих Даних у правоохоронній діяльності.	32	6	4			22	
Всього за семестр № 1:	180	28	16	16		120	екзамен
Всього по дисципліні	180	28	16	16		120	

2. Методичні вказівки до семінарських занять

Тема № 1. Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері.

Семінарське заняття 1. Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері.

Навчальна мета заняття: сформувати у студентів уявлення про сучасні тенденції у сфері інформаційних технологій, зміст задач аналізу Великих Даних та їх вплив на правоохоронну сферу.

Час проведення: 2 год.

Навчальні питання:

1. Третя технологічна революція.
2. Четверта технологічна революція.
3. Використання досягнень Четвертої технологічної революції кримінальними угрупованнями.
4. Вплив на правоохоронну сферу.

Теми для рефератів:

1. Драйвери Четвертої технологічної революції.
2. Використання Штучного інтелекту ОЗУ.
3. Використання дронів ОЗУ.
4. Використання роботів ОЗУ.
5. Використання біотехнологій ОЗУ.
6. Топ-10 стратегічних трендів розвитку технологій у 2020 році.
7. Топ-10 стратегічних трендів розвитку технологій у 2021 році.

Література:

Основна.

1. Інформаційні технології у правоохоронній діяльності. Частина 1: Високотехнологічні тренди у правоохоронній сфері зарубіжних країн: навч. посіб. / Харків. Нац. Ун-т внутр. Справ; [В.М. Струков, Д.В. Узлов, Ю.В. Гнусов та ін.] ; за заг. ред. канд. техн. наук, доц. В.М. Струкова. Харків : ТОВ «ДІСА ПЛЮС», 2020. 276 с.
2. Конспект лекцій.

Додаткова.

3. Зацеркляний М.М. Інформаційні технології у правозастосовній діяльності: Навч. посібник / М.М. Зацеркляний, В.М. Струков. : Х.: ТОВ „Східно-регіональний центр гуманітарно-освітніх ініціатив”; 2010. 332 с.
4. Зацеркляний М.М. Основи комп'ютерних технологій для економістів: Навч. посібник / М.М. Зацеркляний, О.Ф. Мельников, В.М. Струков. – К.: ВД „Професіонал”, 2006 р. – 672 с.

Інформаційні ресурси в Інтернеті.

5. Gartner: Топ-10 стратегічних трендів розвитку технологій у 2020 році: сайт. URL: <https://ain.ua/2018/10/26/gartner-top-10-trendov-razvitiya-texnologij/> (дата звернення: 25.10.2019).

Тема № 1. Актуальність і тенденції аналізу Великих Даних у правоохоронній сфері.

Семінарське заняття 2. Використання аналізу Великих Даних у правоохоронній сфері: досвід зарубіжних країн.

Навчальна мета заняття: сформулювати у студентів уявлення про сучасні тенденції у сфері використання аналізу Великих Даних у правоохоронній сфері (знайомство з досвідом зарубіжних країн).

Час проведення: 2 год.

Навчальні питання:

1. Прогнозування злочинності на основі Big Data.
2. Великі Дані у протидії кіберзлочинності.
3. Протидія фінансовим шахрайствам.

4. Стан аналізу Великих Даних у діяльності Національної поліції України.

Теми для рефератів:

1. Прогнозування злочинності на основі Big Data: система PredPol.
2. Система Palantir (США).
3. Система I2 (США).
4. Система Maltego.
5. Проект N4G(США).
6. Досвід Великобританії.
7. Досвід Нідерландів.

Література:

Основна.

1. Інформаційні технології у правоохоронній діяльності. Частина 1: Високотехнологічні тренди у правоохоронній сфері зарубіжних країн: навч. посіб. / Харків. Нац. Ун-т внутр. Справ; [В.М. Струков, Д.В. Узлов, Ю.В. Гнусов та ін.] ; за заг. ред. канд. техн. наук, доц. В.М. Струкова. Харків : ТОВ «ДІСА ПЛЮС», 2020. 276 с.

2. Конспект лекцій.

Додаткова.

3. Зацеркляний М.М. Інформаційні технології у правозастосовній діяльності: Навч. посібник / М.М. Зацеркляний, В.М. Струков. : Х.: ТОВ „Східно-регіональний центр гуманітарно-освітніх ініціатив”; 2010. 332 с.

4. Зацеркляний М.М. Основи комп'ютерних технологій для економістів: Навч. посібник / М.М. Зацеркляний, О.Ф. Мельников, В.М. Струков. – К.: ВД „Професіонал”, 2006 р. – 672 с.

Інформаційні ресурси в Інтернеті.

5. Gartner: Топ-10 стратегічних трендів розвитку технологій у 2020 році: сайт. URL: <https://ain.ua/2018/10/26/gartner-top-10-trendov-razvitiya-texnologij/> (дата звернення: 25.10.2019).

Тема № 2. Задачі і етапи опрацювання Великих Даних.

Семінарське заняття 3. Задачі обробки Великих Даних та їх особливості.

Навчальна мета заняття: сформулювати у студентів уявлення про сутність задач обробки Великих Даних.

Час проведення: 2 год.

Навчальні питання:

1. Задачі обробки Великих Даних: класифікація, регресія, асоціація, кластеризація, аналіз викидів.
2. Специфіка проблем, що виникають під час обробки Великих Даних.
3. Етапи опрацювання Великих Даних. Препроцесінг даних, аналітична

обробка даних.

4. Типи оброблюваних даних: числові, категоріальні, рангові, текстові, змішані.

5. Оцінювання близькості точок даних у просторі. Масштабування та нормування даних.

Теми для рефератів:

1. Задача класифікації.
2. Задача кластеризації.
3. Задача регресії.
4. Задача асоціації.
5. Метрики у просторі багатовимірних даних.
6. Методи масштабування даних.
7. Методи нормування даних.
8. Препроцесінг даних.
9. Типи оброблюваних даних.

Література:

Основна.

1. Aggarwal C.C. Data Mining. – Cham: Springer Ltd. Publ. Switzerland, 2015. – 734p.
2. Конспект лекцій.

Додаткова.

3. Han L., Kamber M. Data Mining: Concepts and Techniques. – Amsterdam: Morgan Kaufman Publ., 2006. – 754p.
4. Westphal C. Data Mining for Intelligence, Fraud and Criminal Detection. Advanced Analytic & Information Sharing Technologies / C. Westphal. – Boca Raton : CRC Press, 2009. – 426p.
5. Mena J. Investigative Data Mining for Security and Criminal Detection. – Amsterdam: Elsevier Science, 2003. – 452p.

Інформаційні ресурси в Інтернеті.

6. Методи і засоби аналізу даних [Електронний ресурс]. – Режим доступу: <https://bourabai.ru/tpoi/analysis6.htm>.
7. Biryukov A. Neural network clustering methods to evaluate the totality of taxpayers according to their degree of creditworthiness // Artificial societies. 2017. V. 12. Issue 1-2 [Electronic resource]. Access for registered users. URL: [http://artsoc.jes.su/s207751800000103-2-1\(circulationdate:22.07.2018\).DOI:10.18254/S0000103-2-1](http://artsoc.jes.su/s207751800000103-2-1(circulationdate:22.07.2018).DOI:10.18254/S0000103-2-1)
8. Кластерний аналіз [Електронний ресурс]. – Режим доступу: <http://statsoft.ru/home/textbook/modules/stcluan.html>

Тема № 3. Елементи теорії множин і теорії графів.

Семінарське заняття 4. Елементи теорії множин і теорії графів.

Навчальна мета заняття: сформувати у студентів основи теоретичних знань з теорії множин і теорії графів.

Час проведення: 2 год.

Навчальні питання:

1. Поняття множини. Способи задання множин. Властивості множин.
2. Операції над множинами.
3. Поняття графа. Види графів.
4. Способи опису графів. Планарність графів.
5. Алгоритми на графах. Найкоротший шлях, найкоротший гамільтоновий цикл, найкоротше остовне дерево.

Теми для рефератів:

1. Операції над множинами.
2. Алгоритм пошуку найкоротшого шляху між двома вершинами.
3. Алгоритм пошуку найкоротшого гамільтонового циклу.
4. Алгоритм Прима (Краскала).
5. Максимально зв'язні компоненти.
6. Планарність графів.
7. Ізоморфізм графів.

Література:

Основна література.

1. Капітонова Ю.В., Кривий С.Л., Летичевський О.А., Луцький Г.М., Печорін М.К. „Основи дискретної математики”, -Київ.: Наук.думка, 2002. – 579 с.
2. Конспект лекцій.

Інформаційні ресурси в Інтернеті.

3. Основні поняття теорії графів: сайт. URL:
https://evgavrilenko.ucoz.ru/DS/LEKCIYA_1.pdf

Тема № 5. Алгоритми ієрархічної кластеризації.

Семінарське заняття 5. Кластеризація: ієрархічні алгоритми.

Навчальна мета заняття: сформувати у студентів основи теоретичних знань з ієрархічних алгоритмів кластеризації.

Час проведення: 2 год.

Навчальні питання:

1. Види ієрархічних алгоритмів кластеризації.
2. Агломеративні алгоритми кластеризації.
3. Дивизимні алгоритми кластеризації.

4. Методика побудови дендрограм.
5. Міжкластерна відстань. Методи визначення.
6. Види центрів кластерів. Способи визначення координат центрів кластерів.

Теми для рефератів:

1. Агломеративні алгоритми кластеризації.
2. Дивизимні алгоритми кластеризації
3. Види міжкластерних відстаней.
4. Види центрів кластерів. Способи визначення координат центрів кластерів.
5. Монотетичні і політетичні алгоритми дивизимної кластеризації.

Література:

Основна.

1. Aggarwal C.C. Data Mining. – Cham: Springer Ltd. Publ. Switzerland, 2015. – 734p.
2. Aggarwal C.C., Reddy C.K. Data Clustering. Algorithms and Applications.- New York: CRC Press, Taylor & Francis Group, 2014. – 648p.
3. Конспект лекцій.

Інформаційні ресурси в Інтернеті.

4. Методи і засоби аналізу даних [Електронний ресурс]. – Режим доступу: <https://bourabai.ru/tpoi/analysis6.htm>.
5. Biryukov A. Neural network clustering methods to evaluate the totality of taxpayers according to their degree of creditworthiness // Artificial societies. 2017. V. 12. Issue 1-2 [Electronic resource]. Access for registered users. URL: [http://artsoc.jes.su/s207751800000103-2-1\(circulationdate:22.07.2018\).DOI:10.18254/S0000103-2-1](http://artsoc.jes.su/s207751800000103-2-1(circulationdate:22.07.2018).DOI:10.18254/S0000103-2-1)
6. Кластерний аналіз [Електронний ресурс]. – Режим доступу: <http://statsoft.ru/home/textbook/modules/stcluan.html>

Тема № 6. Методи і алгоритми чіткої кластеризації.

Семінарське заняття 6. Алгоритми чіткої кластеризації.

. Навчальна мета заняття: сформувати у студентів основи теоретичних знань з алгоритмів чіткої кластеризації.

Час проведення: 2 год.

Навчальні питання:

1. Класифікація алгоритмів чіткої кластеризації.
2. Алгоритми кластеризації k-means.
3. Алгоритми c-means.
4. Графові алгоритми.
5. Генетичні алгоритми.

Теми для рефератів:

1. Проблема вибору першопочаткових центрів в алгоритмах k-means.
2. Група алгоритмів кластеризації k-means, етапи і особливості.
3. Група алгоритмів кластеризації c-means, етапи і особливості.
4. Графові алгоритми: алгоритм виділення зв'язаних компонент.
5. Графові алгоритми: алгоритм мінімального покриваючого дерева.
6. Графові алгоритми: алгоритм пошарової кластеризації.
7. Генетичні алгоритми.

Література:**Основна.**

1. Aggarwal C.C. Data Mining. – Cham: Springer Ltd. Publ. Switzerland, 2015. – 734p.
2. Aggarwal C.C., Reddy C.K. Data Clustering. Algorithms and Applications.- New York: CRC Press, Taylor & Francis Group, 2014. – 648p.
3. Конспект лекцій.

Інформаційні ресурси в Інтернеті.

4. Методи і засоби аналізу даних [Електронний ресурс]. – Режим доступу: <https://bourabai.ru/tpoi/analysis6.htm>.
5. Biryukov A. Neural network clustering methods to evaluate the totality of taxpayers according to their degree of creditworthiness // Artificial societies. 2017. V. 12. Issue 1-2 [Electronic resource]. Access for registered users. URL: [http://artsoc.jes.su/s207751800000103-2-1\(circulationdate:22.07.2018\).DOI:10.18254/S0000103-2-1](http://artsoc.jes.su/s207751800000103-2-1(circulationdate:22.07.2018).DOI:10.18254/S0000103-2-1)
6. Кластерний аналіз [Електронний ресурс]. – Режим доступу: <http://statsoft.ru/home/textbook/modules/stcluan.html>

Тема № 7. Інструментальні засоби обробки Великих Даних в поліцейській діяльності.

Семінарське заняття 7. Система ePOOLICE раннього виявлення загроз ОЗУ.

Навчальна мета заняття: сформувати у студентів теоретичні знання та уявлення про задачі, функції та можливості системи ePOOLICE раннього виявлення загроз ОЗУ.

Час проведення: 4 год.

Навчальні питання:

1. Загальна характеристика системи ePOOLICE.
2. Сканування відкритих даних в системі ePOOLICE.
3. Розпізнавання загроз ОЗ в системі ePOOLICE.
4. Індикатори аналізу ОЗ в системі ePOOLICE.
5. «Слабкі сигнали» для прогнозу ОЗ в системі ePOOLICE.
6. Якісні методи аналізу та прогнозування ОЗ в системі ePOOLICE.
7. Великі Дані, персональні дані і права людини в системі ePOOLICE.
8. Поліцейська розвідка за відкритими джерелами в системі ePOOLICE.

Теми для рефератів:

1. Джерела інформації в системі ePOOLICE.
2. Сканування соціальних мереж в системі ePOOLICE..
3. Система індикаторів в системі ePOOLICE.
4. Прогнозування ОЗ в системі ePOOLICE.
5. Проблема гармонізації індивідуального права з громадськими, державними і міжнародними правами в системі ePOOLICE.
6. Застосування технологій Великих даних в системі ePOOLICE.
7. Засоби розпізнавання загроз ОЗ в системі ePOOLICE.

Література:

Основна.

1. Струков В.М., Узлов Д.В., Гнусов Ю.В., Колісник Т.П., Можаяев М.О. Інформаційні технології у правоохоронній діяльності. Частина 1: Високотехнологічні тренди у правоохоронній сфері зарубіжних країн. Навчальний посібник. Харків: ХНУВС, 2020. 256 с.
2. Конспект лекцій.

Додаткова.

3. Aggarwal C.C. Data Mining. – Cham: Springer Ltd. Publ. Switzerland, 2015. – 734p.

Інформаційні ресурси в Інтернеті.

4. Методи і засоби аналізу даних [Електронний ресурс]. – Режим доступу: <https://bourabai.ru/tpoi/analysis6.htm>.
5. Biryukov A. Neural network clustering methods to evaluate the totality of taxpayers according to their degree of creditworthiness // Artificial societies. 2017. V. 12. Issue 1-2 [Electronic resource]. Access for registered users. URL: [http://artsoc.jes.su/s207751800000103-2-1\(circulationdate:22.07.2018\).DOI:10.18254/S0000103-2-1](http://artsoc.jes.su/s207751800000103-2-1(circulationdate:22.07.2018).DOI:10.18254/S0000103-2-1)

6. Кластерний аналіз [Електронний ресурс]. – Режим доступу:
<http://statsoft.ru/home/textbook/modules/stcluan.html>