

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»

Затверджено
Вченою радою ННК «ПСА»
Протокол від 28.02.2017 № 2



Заступник голови

В.Д.Романенко

2017 р.

ПРОГРАМА

атестаційного (фахового) випробування
на навчання за освітньо-професійною програмою підготовки «Магістр»
спеціальність 122 «Комп'ютерні науки та інформаційні технології»
спеціалізація «Інтелектуальний аналіз даних і управлінні проектами»

Програму рекомендовано
атестаційною комісією
кафедри математичних методів
системного аналізу

Протокол засідання кафедри
від 15.02.2017 р. № 5

В. о. завідувача кафедри

 О.Л.Тимошук
«15» 02 2017 р.

Київ
2017

ВСТУП

Програма вступного комплексного фахового випробування на навчання за освітньо-професійною програмою підготовки «Магістр» за спеціальністю 122 Комп'ютерні науки та інформаційні технології (*спеціалізація* Інтелектуальний аналіз даних в управлінні проектами) (далі — Програма) призначена для отримання досвіду самостійної роботи абітурієнта з підготовки до екзамену.

Метою програми є формування у абітурієнтів здатності ознайомитися із предметними питаннями курсів навчальних дисциплін, що включені в екзаменаційні білети; опрацювати підручники, навчальні посібники та інші інформаційно-літературні джерела предметної області знання; осмислено упорядочити і систематизувати засвоєні теоретичні знання і практичні навички; вмотивовано виконати роботу на екзамені, продемонструвавши певний рівень засвоєння навчальних дисциплін в результаті навчання.

Перелік навчальних дисциплін цієї Програми складають такі, що відповідно належать до циклу професійної та практичної підготовки навчального плану підготовки «бакалавра» 6.050101 «Комп'ютерні науки»:

- 1) Прикладна статистика
- 2) Дослідження операцій
- 3) Бази даних та інформаційні системи

Вступний екзамен проводиться чотири академічних години без перерви (180 хвилин), в аудиторному приміщенні випускової кафедри, методом одержання екзаменаційного білету—повернення письмової роботи. Завданням на екзамені є розв'язання завдань екзаменаційного білету. Екзаменаційний білет містить п'ять практичних за типом завдань. Диференціації робочого часу, відведеного на виконання кожного завдання, немає. Фіксується час початку і закінчення роботи.

ПЕРЕЛІК НАВЧАЛЬНОГО МАТЕРІАЛУ, ЩО ВІНОСИТЬСЯ НА ФАХОВЕ ВИПРОБУВАННЯ

1 Прикладна статистика

Два значення слова «статистика». Чотири етапи статистичного аналізу даних. Два типи даних: часові ряди і часові перерізи; два методи збору статистичних даних: суцільний і вибірковий. Планування експерименту і збір даних. Попередня обробка та дослідження (тестування) даних. Методи заповнення пропусків у часових рядах, нормування даних, виявлення екстремальних значень. Цифрова і оптимальна фільтрація даних.

Поняття детермінованого і випадкового процесу. Випадкові процеси з детермінованими складовими. Формулювання гіпотез стосовно типів розподілів ймовірностей, можливих ситуацій. Деякі типи розподілів

дискретних випадкових величин: нормальний, розподіл Пірсона (χ^2), t – розподіл Стюдента, експоненціальний, Вейбула.

Математичні і статистичні моделі. Поняття структури математичної моделі. Стаціонарні та нестаціонарні процеси, тестування фактичних даних (тести на інтегрованість, гетероскедастичність, лінійність). Встановлення належності даних до дискретного розподілу конкретного типу: коефіцієнт асиметрії, ексцес, статистика Жак-Бера, статистика χ^2 , критерії Колмогорова-Смірнова та Андерсона-Дарлінга.

Загальна методика перевірки гіпотез. Нульова і альтернативна гіпотези. Поняття рівня значущості. Помилки першого і другого роду. Критерій перевірки нульової гіпотези. Критична область, потужність критерію, область прийняття гіпотези. Перевірка гіпотези стосовно значущості оцінок параметрів моделей.

Задачі оцінювання структури і параметрів статистичних моделей, формулювання гіпотез стосовно значущості оцінок та якості (адекватності) статистичних моделей. Перевірка сформульованих гіпотез стосовно оцінок параметрів моделей, прийняття рішення стосовно адекватності моделі.

Кореляційний аналіз даних. Неперервні і дискретні кореляційні функції. Теоретична автокореляційна функція (АКФ), рівняння Юла-Уокера. Автокореляційна і часткова автокореляційна функції, функція взаємної кореляції двох величин. Кореляційна матриця.

Регресійний аналіз. Типи регресійних рівнянь. Вимоги до даних, оцінок параметрів моделі та до моделі у цілому. Методика побудови регресійної моделі. Адекватність побудованої моделі, критерії адекватності моделі.

Метод найменших квадратів (МНК), властивості оцінок, обчислених за МНК. Рекурсивний метод найменших квадратів (РМНК). Відмінності рекурсивного МНК від звичайного МНК. Можливі випадки застосування РМНК. Порівняння властивостей оцінок, отриманих за МНК, з властивостями оцінок, обчислених за РМНК.

Оцінювання параметрів за методами максимальної правдоподібності та Монте Карло для марковських ланцюгів.

Метод головних компонент: призначення, алгоритм реалізації. Нелінійний метод головних компонент.

Суть та призначення байєсівських методів статистичної обробки даних (БАД). Недоліки та переваги БАД перед класичним підходом. Приклади застосування.

Список літератури [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

2 Дослідження операцій

ЛІНІЙНЕ ПРОГРАМУВАННЯ (ЛП)

Метод розв'язання задач ЛП з довільним видом обмежень, оснований на штучних змінних.

Двоїста задача ЛП. Двоїстий симплекс-метод.

Метод оберненої матриці.

Дослідження меделей ЛП-задач на чутливість.

Транспортні задачі. Метод потенціалів.

ДИСКРЕТНЕ ПРОГРАМУВАННЯ(ДП)

Метод відсікаючих площин Гоморі.

Метод гілок та меж.

Метод гілок та меж в задачі комівояжера.

Метод послідовного аналізу та відсіву варіантів (ПАВ) в задачі ЛЦП.

Метод послідовного аналізу та відсіву варіантів (ПАВ) в задачі булево програмування.

НЕЛІНІЙНЕ ПРОГРАМУВАННЯ

Метод множників Лагранжа.

Задача квадратичного програмування. Умови оптимальності Куна-Таккера для задач квадратичного програмування.

Метод розв'язання загальної задачі геометричного програмування ГП з ступенем складності $d > 0$.

Методи можливих напрямків. Метод Зойтендейка у випадку лінійних обмежень.

Методи можливих напрямків. Метод Зойтендейка у випадку нелінійних обмежень-нерівностей.

Прямі методи пошуку.

ЗАДАЧІ БАГАТОКРИТЕРІАЛЬНОЇ ОПТИМІЗАЦІЇ

Багатокритеріальна задача. Метод обмежень.

Багатокритеріальний вибір альтернатив на основі нечіткого відношення переваги.

Список літератури [11, 12, 13, 14, 15, 16, 17]

3 Бази даних та інформаційні системи

ОСНОВНІ ПОНЯТТЯ. Інформація, дані, знання. Аспекти роботи з даними. Поняття про інформаційні технології (ІТ). Особливості та завдання інформаційних систем (ІС). Файл, запис, поле. Файлові інформаційні системи (ФІС). Ідея СКБД, відміни від ФІС. Визначення банку даних (БнД). Вимоги до БнД. Переваги централізації керування даними.

ТЕХНОЛОГІЇ СТВОРЕННЯ ІС. Життєвий цикл інженерного виробу. Моделі життєвого циклу розробки ІС. Задачна модель. Каскадна модель. Спіральна модель. Загальна технологія створення ІС та АС. Основи побудови банків даних БнД.

АНАЛІЗ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПІДПРИЄМСТВА. Підприємство як відкрита система. Метаболізм підприємства. Моделювання за допомогою діаграм потоків даних і подій (Data Flow Diagrams). Матриці подій.

МЕТОДОЛОГІЯ СТРУКТУРНОГО АНАЛІЗУ ТА ПРОЕКТУВАННЯ SADT. Історичний розвиток технології SADT. Склад функціональної моделі.

Ієрархія діаграм. Типи зв'язків між функціями. Міжнародні стандарти системи IDEF.

МОДЕЛІ ДАНИХ. Поняття моделі даних (МД). Сильно- та слабкоструктуровані МД. Структури, операції, обмеження МД. Модель «сутність-зв'язок». Типи зв'язків. Степені зв'язку, залежність по коду. N-зв'язки. Композиція зв'язків. Типи і підтипи (ролі). Поняття життєвого циклу об'єкта (екземпляр сутності). Початок, закінчення, координація ЖЦ. Обмеження цілісності, бізнес-правила. Локальні інфологічні моделі. Побудова глобальної інфологічної моделі.

РЕЛЯЦІЙНА МД. Реляційні алгебра та числення. Базові поняття реляційних баз даних. Реляційна модель даних. Реляційна алгебра та її операції. Реляційне числення на кортежах. Реляційне числення на доменах.

НОРМАЛІЗАЦІЯ БД. Групування атрибутів у відношення. Аномалії та їхні види. 1НФ, 2НФ, 3НФ, BCNF, 4НФ, 5НФ.

ДОРЕЛЯЦІЙНІ МД. Ієрархічна МД. Мережева МД.

БАНК ДАНИХ І ЙОГО МІСЦЕ В АВТОМАТИЗОВАНИХ СИСТЕМАХ. Визначення банку даних. (БнД). Вимоги до БнД. БнД як автоматизована система. Види забезпечення. Архітектура БнД. Адміністратор БД і його функції. Довідник даних. Дво- та трирівневі архітектури БнД. Централізація і децентралізація процесів обробки даних.

МОВА БД SQL. Мови опису даних і маніпулювання даними. Реляційні операції як команди мови маніпулювання даними. Віртуальні атрибути і таблиці. Приклади використання операторів Insert, Update Deleteю

ЧАСТИНА FROM КОМАНДИ SELECT. Структура запиту. Частина FROM команди SELECT.

ОСОБЛИВОСТІ КОМАНДИ SELECT. Зміна порядку виведення рядків (ORDER BY). Усунення дублювання (модифікатор DISTINCT). З'єднання (JOIN). Використання агрегатних функцій з угрупованням. Порядок виконання оператора SELECT.

ВИКОРИСТАННЯ SQL В ПРИКЛАДНИХ ПРОГРАМАХ. Тригери та цілісність посилення. Збережені процедури. Використання курсорів. Динамічний SQL.

СУЧАСНІ ТЕНДЕНЦІЇ РОЗВИТКУ ТЕХНОЛОГІЙ В СКБД. Економічні тенденції. Технічні тенденції.

ВИБІР СУЧАСНОЇ СКБД. Технічні тенденції. Порівняння характеристик СКБД. Особливості сегментування СКБД. Вартість пропрієтарних СКБД.

ПАРАЛЕЛЬНІ БАЗИ ДАНИХ. Основні поняття паралельної обробки даних. Архітектура багатопроцесорних систем. Розподіл даних. Паралельна обробка запитів.

ДЕДУКТИВНІ БАЗИ ДАНИХ. Основні поняття дедуктивних баз даних. Інтерпретація логічних правил. Мова Datalog. Обчислення нерекурсивних Datalog-програм. Обчислення рекурсивних програм. Обчислення правил із запереченням.

БАЗИ ДАНИХ В ІНТЕРНЕТІ. Основи XML. Базы даних на основі XML. Базы даних із вбудованою підтримкою XML. XML-БД на основі баз даних іншого типу. Мови запитів. Робота з базами даних через мережу Інтернет.

CASE-ЗАСОБИ АВТОМАТИЗАЦІЇ ПРОЕКТУВАННЯ БД. PowerDesigner. ERwin.

ПРОЕКТУВАННЯ КОРИСТУВАЧЬКОГО ІНТЕРФЕЙСУ. Usability, значення і міфи. Проблеми проектування інтерфейсів користувача (ІК). Вимоги до ІК. Принципи реалізації інтерфейсу. Етапи проектування ІК. Методи і критерії оцінки ІК. Основні відомості з інженерної психології. Приклад стандарту ІК від компанії IBM.

ФІЗИЧНА ОРГАНІЗАЦІЯ ІС. Структура зовнішньої пам'яті. Зберігання таблиць. Індеси та В-дерева. Хеш-функція. Службова інформація.

СИСТЕМА «КЛІЄНТ—СЕРВЕР». Історичний розвиток обчислень в автоматизованих системах. Модель файлового серверу. Модель віддаленого доступу до даних. Модель сервера бази даних. Модель сервера додатків. Відкриті системи.

ТРАНЗАКЦІЇ. Що таке транзакція. Транзакції і цілісність БД. Рівні ізольованості транзакцій (РІТ). Серіалізація транзакцій. Конфлікти між транзакціями. Синхронізаційні захоплення. Гранульовані синхронізаційні захоплення. Предикатні синхронізаційні захоплення. Поняття глухого кута (Dead End). Граф очікування транзакцій. Метод тимчасових міток.

ТАБЛИЦІ ТА УЯВЛЕННЯ. Типи таблиць. Таблиці, що організовані у вигляді кучі. Таблиці, що організовані по індексу. Таблиці в індексному кластері. Таблиці в хеш-кластері. Вкладені таблиці. Тимчасові таблиці. Об'єктні таблиці. Уявлення. Використання уявлень в SQL-запитах.

ІНДЕКСИ НА ОСНОВІ В-ДЕРЕВА. Індеси на основі бітових карт. Індеси по функціях. Прикладні індеси.

РОЗПОДІЛЕНІ БАЗИ ДАНИХ. Розподілені БД. Причини виникнення і завдання. Розподіл даних по мережі. Проблеми розподілених БД. Види обмежень цілісності у розподілених ІС.

ПРОБЛЕМИ ЗАХИСТУ ТА БЕЗПЕКИ ДАНИХ. Захист, безпека та секретність. Обов'язки адміністратора БД із захисту даних. Ідентифікація користувача в БД. Вибірче і обов'язкове керування доступом. Базові засоби аутентифікації при цифровій передачі даних.

СИСТЕМИ АНАЛІТИЧНОЇ ОБРОБКИ ДАНИХ OLAP. Оперативна OLAP та аналітична OLAP обробки даних. Концепція сховища даних. Властивості інформаційних сховищ. Типи OLAP. Основні компоненти інформаційного сховища. Робота з гіперкубом. Реалізація сховищ і вітрин даних. Проблеми інтеграції даних. Наявні програмні реалізації.

ІНФОРМАЦІЙНО-ПОШУКОВІ СИСТЕМИ (ІПС). Інформаційно-пошукові системи і їхні відмінності від банків даних. Пошукові структури в ІПС. Оцінка ефективності ІПС.

ОБ'ЄКТНО-ОРІЄНТОВАНІ БАЗИ ДАНИХ. Об'єктно-орієнтована модель ODMG. Мова опису об'єктів ODL ODMG. Об'єктна мова запитів OQL ODMG. Архітектура OOCKБД. Зображення об'єктної моделі в реляційній базі даних.

БАЗИ ЗНАНЬ. Відколи дані стають знаннями. Постулати системи баз даних. Моделі зображення знань. Розширення семантики даних. Нечіткі дані. Механізми виведення даних.

ЖУРНАЛІЗАЦІЯ ЗМІН БД. Журналізація і буферизація. Індивідуальний відкат транзакції. Відновлення після м'якого збою. Фізична узгодженість бази даних. Відновлення після жорсткого збою.

ПРОБЛЕМИ ОПТИМІЗАЦІЇ SQL-ЗАПИТІВ. Загальна схема обробки запиту. Синтаксична оптимізація запитів. Семантична оптимізація запитів. Вибір та оцінка альтернативних планів виконання запитів.

ВІДСУТНЯ ІНФОРМАЦІЯ. Огляд концепції тризначної логіки. Наслідки викладеної схеми. Відсутні значення і ключі. Зовнішні з'єднання. Спеціальні значення. Засоби мови SQL.

ПІДТРИМКА ПРИЙНЯТТЯ РІШЕНЬ. Особливості технології підтримки прийняття рішень. Проектування баз даних для підтримки прийняття рішень. Підготовка даних. Сховища даних і магазини даних. Оперативна аналітична обробка. Розробка даних. Засоби SQL.

НАСЛІДУВАННЯ ТИПІВ. Ієрархія типів. Поліморфізм та змінність. Змінні та оператори присвоювання. Уточнення за допомогою обмежень. Операції порівняння. Оператори, версії та сигнатури. Аналіз взаємодії між типами і підтипами. Додаткова інформація про уточнення за допомогою обмежень. Засоби мови SQL.

Список літератури [18, 19, 20, 21, 22, 23, 24, 25, 26]

ПРИКІНЦЕВІ ПОЛОЖЕННЯ

Користування допоміжним матеріалом на екзамені

— Забороняється

Критерії оцінювання (за системою ECTS, стобальна шкала)

Розв'язання кожної задачі оцінюється за такими критеріями:

95—100	—	задачу розв'язано повністю, вірно
85—94	—	задачу розв'язано вірно, відповідь правильна, але наявними є один-два недоліки (наявними є деякі методичні помилки, порушено послідовність викладок тощо)
75—84	—	задачу розв'язано вірно, але відповідь неправильна (наявними є арифметичні помилки)
65—74	—	задачу розв'язано неповністю, але намічено правильний хід розв'язування
60—64	—	задачу не розв'язано, але наведено формули або твердження, що можуть бути використані при розв'язуванні задачі
менше 60	—	задачу не розв'язано

Результат роботи обчислюється як середнє арифметичне оцінок, що їх отримано за кожну задачу і заокруглюється до цілих.

Приклад типового завдання комплексного фахового випробування

Білет № 0

1. Із великої партії мікросхем підсилювального типу випадковим чином відібрали і перевірили 100 штук. Коефіцієнт підсилення кожної мікросхеми повинен бути не меншим 10. У 36-и мікросхем він виявився меншим 10.

а) Необхідно знайти 95%-й довірчий інтервал для частки таких мікросхем у всій партії.

б) Який мінімальний обсяг вибірки потрібно виготовити для того, щоб із ймовірністю 0,95 можна було стверджувати, що частка мікросхем, коефіцієнт підсилення яких менший 10, у всій партії відрізняється від частоти появи таких мікросхем у вибірці не більше ніж на 1%?

2. Визначити оптимальне рішення задачі дискретного програмування методом ПАВ

$$F = 5x_1 - 3x_2 \rightarrow \min$$

$$3x_1 + 2x_2 \geq 6$$

$$2x_1 - 3x_2 \geq -6$$

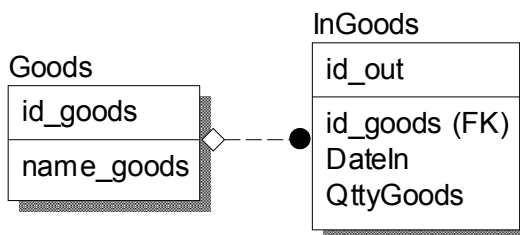
$$x_1 - x_2 \leq 4$$

$$4x_1 + 7x_2 \leq 28$$

$$x_1, x_2 \geq 0$$

$$x_1 \in Z, x_2 \in Z$$

3. Відповідно до заданої схеми написати запит до бази даних для обчислення сумарної кількості товарів, що надійшли до магазину з конкретної дати D0 до конкретної дати D1 включно, тобто $DateIn \geq D0$ і $DateIn \leq D1$. Таблиця Goods містить перелік товарів. В таблиці InGoods фіксується надходження товарів в магазин, в полі DateIn зберігається дата надходження, в полі QtyGoods — кількість товарів, що надійшли. Результат повинен мати вигляд: назва товару, розрахована сума.



СПИСОК ЛІТЕРАТУРИ

1. Бідюк П.І. Прикладна статистика / П. І. Бідюк, О.М. Терентьєв, Т.І. Просянкіна-Жарова. — Вінниця: «Едельвейс і К», 2013. — 288 с.
2. Мелник М. Основы прикладной статистики / Мелник М. — М.: Энергоатомиздат, 1983. — 415 с.
3. Сигел Э.Ф. Практическая бизнес-статистика / Э.Ф. Сигел. — СПб: Виьямс, 2002. — 1052 с.
4. Сеньо П.С. Теорія ймовірностей та математична статистика / С. С. Сеньо. — Київ: Центр навчальної літератури, 2004. — 448 с.
5. Gelman A. Bayesian Data Analysis / A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin. — New York: Chapman & Hall, 2000. — 670 с.
6. Чернов Г. Элементарная теория статистических решений / Г. Чернов, Л. Мозес. — М.: «Советское Радио», 1982. — 407 с.
7. Хей Дж. Введение в методы байесовского статистического вывода / Дж. Хей. — М.: «Финансы и статистика», 1987. — 336 с.
8. Андерсон Т. Статистический анализ временных рядов / Т. Андерсон. — М.: Мир, 1976. — 755 с.
9. Бідюк П.І. Моделі і методи математичної статистики / П. І. Бідюк. — Київ: НТУУ «КПІ» (електронне видання), 2014. — 710 с.
10. Casella G. Statistical Inference / G. Casella, R.L. Berger. — Duxbury (USA): Thomson Learning, 2002. — 660 p.
11. Зайченко Ю. П. Дослідження операцій / Ю. П. Зайченко. — К.: Слово, 2001. — 688 с.
12. Лэдсон Л. Оптимизация больших систем / Л. Лэдсон. — М.: Наука. — 431 с.
13. Михайлович В.С. Методы последовательной оптимизации в дискретных сетевых задачах оптимального распределения ресурсов / В. С. Михайлович, А. И. Кукса. — М.: Наука. — 208 с.
14. Зайченко Ю. П. Исследование операций / Ю. П. Зайченко. — К.: Вища школа. — 552 с.
15. Даффин Р. Геометрическое программирование / Р. Даффин. — М.: Мир. — 308 с.
16. Хедли Дж. Нелинейное и динамическое программирование / Дж. Хедли. — М.: Мир. — 352 с.
17. Навчально-методичний посібник до практичних занять з курсу «Математичні методи оптимізації» для студентів магістратури усіх спеціальностей / Уклад. О.Ю.Зайченко. — К.: Політехніка. — 88 с.
18. Хомоненко А.Д. Базі даних: учебник для вызов по техн. и экон. спец. / А.Д.Хоменко, В.М.Цыганков, М.Г. Мальцев [под ред. А.Д.Хомоненко]. — С-Пб: Корона принт, 2002. — 672 с.
19. Дейт К. Введение в системы баз данных [8-е изд.: пер. с англ.] / Дейт К. — М.: Издательский дом «Вильямс», 2005. — 1328 с.
20. Виейра Р. Программирование баз данных Microsoft SQL Server 2005. Базовый курс [пер. с англ.] / Виейра Р. — М.: Диалектика, 2008. — 832 с.

21 Форта Б. Освой самостоятельно SQL. 10 минут на урок [пер. с англ.] / Форта Б. — М.: Вильямс, 2006. — 288 с.

22 Исаченко А.Н. Модели данных и системы управления базами данных: пособие для студентов / Исаченко А.Н., Бондаренко С.П. — Минск: БГУ, 2007. — 220 с.

23 Грабер М. Введение в SQL / Грабер М. — М.: Лори, 2010. — 228 с.

24 Маклаков С.В. BPwin, ERwin CASE-средства разработки информационных систем / Маклаков С.В. — М.: Диалог-МИФИ, 2001. — 304 с.

25 Пасічник В.В. Організація баз даних та знань / Пасічник В.В., Резніченко В.А. — К.: Видавнича група BVH, 2006. — 384 с.

26 Коннолли Т. Базы данных: проектирование, реализация и сопровождение. Теория, практика [3-е изд.: пер. с англ.] / Коннолли Т., Бегг К. — М.: Издательский дом «Вильямс», 2003. — 1440 с.

Розробники програми:

д.т.н., проф.

д.т.н., доц.

к.т.н.

П.І.Бідюк

О.Ю.Зайченко

Д.Г.Діденко