

РОЗРОБКА ЗАДАЧ ДО ДИСЦИПЛІНИ «АНАЛІЗ ДАНИХ»

Автор роботи:

Ласкавцева Софія Дмитрівна,
учениця 11 класу

Наукові керівники:

Давидов Вячеслав Вадимович,
завідувач кафедри інформаційних технологій
Приватної установи «Університет науки,
підприємництва та технологій», доктор технічних наук

Арзубов Микола Олексійович,
вчитель інформатики



Актуальність

Актуальність полягає у потребі в удосконаленні навчальних програм та методик викладання у вік використання онлайн навчання та онлайн олімпіад із застосуванням сучасних підходів та технологій до вже апробованої системи.

Структура презентації



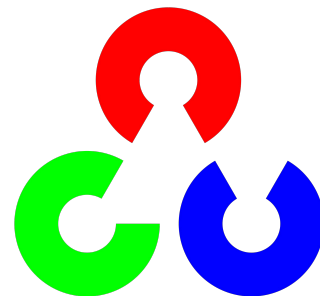
Методологія
розробки задач
для системи
DOTS



NumPy
5 задач



Pandas
5 задач



OpenCV
“Counting objects”

Методологія розробки задач для системи DOTS

DOTS – це тестувальна система для навчання програмуванню, алгоритмам і структурам даних.

Для того, щоб додати задачу в систему DOTS потрібно:

- Умова задачі
- Тести
- Чекер і валідатор

Методологія розробки задач для системи DOTS

Ліміт часу:

Time Limit

Ліміт використання пам'яті:

Memory Limit

Формат вхідних даних

Формат вихідних даних

Обмеження

Приклади

тест	відповідь

Пояснення до прикладів

Зображення умови, згенероване за
пустим шаблоном

```
\begin{problem}{Task Name}{input}{output}{Time Limit}{Memory Limit}

% Тут умова задачі
%
%
%

% Далі треба чітко описати формат вхідних даних
\inputFile

% Далі треба чітко описати формат вихідних даних
\outputFile

% Далі треба вказати обмеження на вхідні дані
\constraints

% Приклади
\examples
\begin{example}
\exmpfile{01.in}{01.out}%
\exmpfile{02.in}{02.out}%
\end{example}

% Пояснення до прикладів (за потреби)
\explanations

\end{problem}
```

Шаблон умови задачі (LaTeX)

Блок-схема алгоритму роботи тестувальної системи DOTS



Терміни

- DataFrame - це двовимірна структура даних в бібліотеці Python Pandas побудована з рядків і стовпців.
- NPZ - це формат файлів від numpy, який забезпечує зберігання масивів даних за допомогою стиснення gzip.
- CSV - текстовий формат, призначений для подання табличних даних. Рядок таблиці відповідає рядку тексту, який містить одне або кілька полів, розділених комами.

Блок задач Pandas

Задача А. “Pandas: sum, max, min”

Обчисліть суму всіх чисел в таблиці, а також знайти максимальний та мінімальний елементи в таблиці.

Навички, що перевіряються: зчитування даних з файлу формату CSV і завантаження їх у DataFrame.

input.csv	відповідь
C1;C2;C3;C4	719
41;55;22;6	100
28;9;13;64	6
60;100;95;52	
13;7;75;79	

Блок задач Pandas

Задача В. “Data Cleaning”

Видаліть рядки, в яких не всі значення є цілими числами та стовпчики, в яких середнє арифметичне всіх значень дорівнює середньому геометричному.

Навички, що перевіряються: видалення стовпців і рядків за певною умовою.

input.csv	Відповідь
5;1;2;3;4;3;5 1;1;2;2;1; ;4 2;3;2;5;4;5;5 1.1;AA;1;2;2;4;6	5;1;3;3 2;3;5;5

Блок задач Pandas

Задача С. “Timestamp”

Файл формату CSV містить такі дані: id рішення, id користувача, id задачі, результат тестування, час відправки в timestamp (POSIX-час, тобто кількість секунд що минула з 00:00:00 UTC 1 січня 1970 року). Треба знайти кількість розв'язків, які пройшли всі тести, та вивести час відправки цих рішень в форматі dd.mm.yyyy hh:mm:ss.

Навички, що перевіряються: TimeStamp (обробка дат і часу).

input.csv	відповідь
sol_id;user_id;task_id;score;posted_time 166142;1017;2004;74.36;1693566723 166143;1038;2006;100.00;1693569290 166145;1024;2007;100.00;1693573289 166148;1085;2002;12.50;1693578174 166152;1032;2002;100.00;1693584082	3 01.09.2023 11:54:50 01.09.2023 13:01:29 01.09.2023 16:01:22

Блок задач Pandas

Задача D. “Statistics”

Файл формату CSV містить такі дані: id рішення, id користувача, id задачі, результат тестування, id мови програмування. Треба зробити статистику, щодо використання мов програмування.

Навички, що перевіряються:
обчислення статистичних показників.

input.csv	Відповідь
<pre>sol_id;user_id;task_id;lang_id;score 153508;1003;2003;3;58.14 153509;1003;2003;2;100.00 153511;1002;2002;4;0.00 153514;1003;2001;4;41.94 153518;1002;2001;4;80.65 153523;1002;2002;4;0.00 153529;1001;2005;5;0.00 153536;1001;2002;3;100.00 153544;1001;2003;1;90.70 153553;1002;2005;1;96.55 153563;1001;2002;5;0.00 153574;1002;2003;1;0.00 153586;1003;2003;5;0.00 153599;1002;2004;3;0.00 153613;1003;2003;5;44.19 153628;1002;2003;1;2.33 153644;1002;2004;1;0.00 153661;1002;2005;1;10.34 153679;1003;2001;2;100.00 153698;1001;2004;5;95.00</pre>	<pre>Pascal: 2 6 0 2 Python: 1 2 2 0 C++: 3 3 1 1 Java: 2 4 0 2 C#: 2 5 0 3</pre>

Блок задач Pandas

Задача Е. “Certification”

Дані з однієї з таблиць бази даних сертифікації зберегли у файлі формату CSV. Цей файл містить таку інформацію: id рішення, id користувача, id задачі, результат тестування. Треба за цими даними вивести результат та вердикт для кожного учасника сертифікації.

Навички, що перевіряються:
обробка і групування даних.

input.csv	Відповідь
sol_id;user_id;task_id;score 159706;1002;2003;29.17 159708;1002;2003;25.00 159711;1002;2002;97.83 159715;1001;2003;62.50 159724;1001;2003;12.50 159730;1002;2002;8.70 159730;1001;2002;30.43 159731;1002;2001;90.00 159739;1002;2001;16.00 159745;1001;2001;50.00 159755;1001;2002;13.04 159763;1001;2003;50.00 159766;1001;2001;26.00 159771;1002;2002;56.52 159781;1001;2001;100.00 159788;1002;2001;32.00 159798;1002;2003;54.17 159801;1001;2002;52.17	1001: Very Good [71.56%] 1002: Excellent [80.67%]

Блок задач NumPy

Задача А. “NumPy: ones, zeroes, full”

Згенеруйте і заповніть квадратичну матрицю розмірністю $N \times N$ нулями. На головній і побічній діагоналях, а також по периметру, замініть нулі на одиниці.

Навички, що перевіряються: заповнення матриць і масивів.

тест	відповідь
5	11111 11011 10101 11011 11111

Блок задач NumPy

Задача В. “NumPy: sum, max, min”

У декількох одновимірних числових NumPy масивах, записаних у файлі формату NPZ, треба знайти: суму всіх елементів; мінімальний елемент; максимальний елемент; індекси всіх мінімальних елементів; індекси всіх максимальних елементів.

Навички, що перевіряються: знаходження індексів елементів, з певним значенням.

input.npz	відповідь
[array([2, -2, -1, 0, 1, -1]), array([1, -3, -3, 3, 2, 2]), array([2, 3, 3, 2, 1, 2, -3])]	-1 -2 2 1 0 2 -3 3 1 2 3 10 -3 3 6 1 2

Блок задач NumPy

Задача С. “NumPy: where”

У декількох одновимірних числових NumPy масивах, записаних у файлі формату NPZ, треба знайти елементи, значення яких знаходиться в діапазоні від 1 до 9 включно.

Навички, що перевіряються: вибір елементів, які задовольняють певну умову.

input.npz	відповідь
<code>[array([10, -12, 1, 5, -1, 2]), array([1, 13, -3, 3, 2, 9]), array([-2, 13, 23, -2, 11, -3])]</code>	3 1 5 2 sum = 8 4 1 3 2 9 sum = 15 0 sum = 0

Блок задач NumPy

Задача D. “Numpy: poly”

Нехай P - поліном n -го ступеня. Відомо, що у цього полінома рівно n дійсних коренів. Потрібно знайти коефіцієнти полінома P ; обчислити значення полінома в заданих точках; обчислити значення похідної від заданого полінома в заданих точках.

Навички, що перевіряються: операції над поліномами.

тест	відповідь
2 1.0 -1.0	1.00000 0.00000 -1.00000 5.25000 8.76562 15.40291 0.00000 8.61620
5 2.5 -3.125 4.05005 -1.0 3.101	5.00000 -6.25000 8.10010 -2.00000 6.20200

Блок задач NumPy

Задача Е. “Numpy: linalg”

Нехай A — прямокутна матриця з N рядків та M стовпців. Відомо що $|M - N| = 2$. Треба написати програму, яка видаляє з матриці A два рядки або два стовпці так, щоб вийшла квадратна матриця B з максимальним детермінантом.

Навички, що перевіряються: обчислення детермінанта матриці; видалення строк і стовпців матриці.

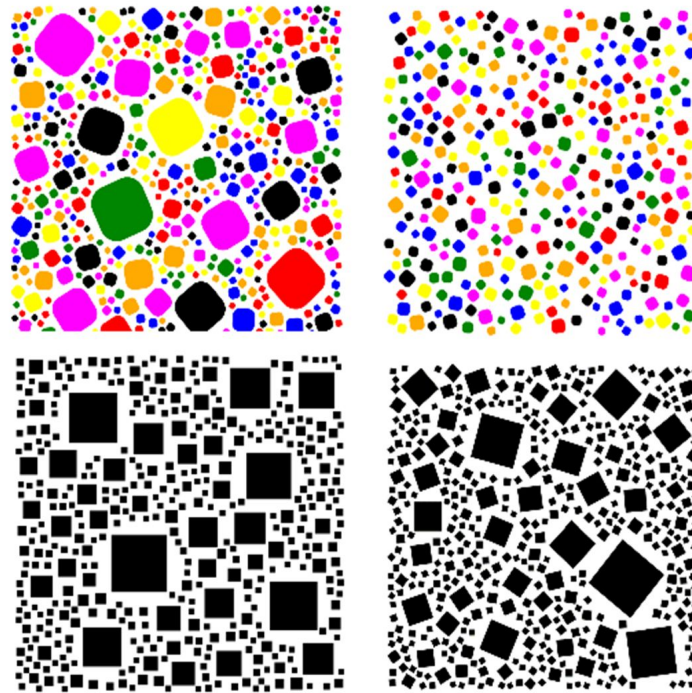
тест	відповідь
1 2 3 4 5 2 1 5 8 9 3 1 4 5 2	3 1 2 5 2 1 9 3 1 2 3 4

Блок задач OpenCV

Задача «Counting Objects»

Потрібно порахувати кількість прямокутників на зображенні формату «PNG». Вони можуть відрізнятися за розміром, поворотом відносно осей координат, кольором (кольорові або чорно-білі також зафарбовані та незафарбовані), та формою кутів (закруглені або гострі).

Для генерації тестів була розроблена програма, яка написана мовою програмування Pascal з використанням бібліотеки GraphABC.



Чотири приклади зображень, що були згенеровані програмою

Висновки

- Розроблено навчальні завдання для системи DOTS, спрямовані на розвиток навичок в аналізі даних та комп'ютерному зорі.

Адаптація задач для тестування в системі DOTS підвищує ефективність тестування знань студентів у цих областях.

- Розроблено шаблони для автоматизованого тестування аналогічних завдань в системі DOTS.

Це дозволяє покращити якість підготовки учнів та студентів у роботі з бібліотеками Python, такими як Pandas, Numpy та OpenCV.

Результати науково-дослідної роботи вже були впроваджені у навчальний процес і доповідались на кафедрі інформатики Харківського національного педагогічного університету імені Г. С. Сковороди.

Вхід до тестувальної системи

Пропоную всім бажаючим спробувати самостійно подивитися результат впровадження роботи в систему DOTS, а також спробувати порозв'язувати задачі за зазначеним посиланням із зазначеними логінами та паролями.

- логін1: man1
пароль1: man1
- логін2: man2
пароль2: man2
- логін3: man3
пароль3: man3



QR Code з посиланням
на сайт системи DOTS

Дякую за увагу!