

ЗАВДАННЯ XXII МІЖНАРОДНОЇ ОЛІМПІАДИ З ІНФОРМАТИКИ ТА РЕКОМЕНДАЦІЇ ЩОДО ЇХ РОЗВ'ЯЗАННЯ

Бондаренко В.В.

(Продовження, початок у №7 за 2010 рік)

ЗАВДАННЯ ПЕРШОГО ТУРУ

3. ЯКІСТЬ ЖИТТЯ

У провінції Альберта міста, зазвичай, сплановані у вигляді прямокутної сітки кварталів. Квартали мають координати від 0 до $(R-1)$ у напрямку «північ-південь» та від 0 до $(C-1)$ у напрямку «захід-схід».

Якість життя у кожному конкретному кварталі оцінено деяким числом, що називається рангом якості, від 1 до $R \times C$, де 1 — найкраща якість, а $R \times C$ — найгірша, причому всі ранги якості є різними.

Відділ планування міста збирається знайти складений з кварталів прямокутник розміру $H \times W$ такий, щоб медіана рангів якості серед усіх кварталів у цьому прямокутнику була найкращою. Тут H — розмір у напрямку «північ-південь», а W — у напрямку «захід-схід». При цьому H і W — непарні числа, що не перебільшують R і C відповідно. *Медіана рангів* якості для непарної кількості рангів якості визначається як таке число m , при якому кількість рангів якості, що кращі за m , дорівнює кількості рангів, що гірші за m .

Ваша задача — написати процедуру `rectangle(R,C,H,W,Q)`, де R і C задають розмір прямокутної сітки міста, H і W задають розмір прямокутника з кварталів, а Q — масив, у якому елемент $Q[a][b]$ — ранг якості для кварталу, що має координату a у напрямку «північ-південь» і координату b — у напрямку «захід-схід».

Ваша реалізація процедури `rectangle` має повернути число, що дорівнює кращій (найменшій) можливій медіані рангів якості серед прямокутників розміру $H \times W$.

У кожному тесті процедура `rectangle` буде запускатися тільки один раз.

Приклад 1

```
R=5, C=5, H=3, W=3,
Q=  5  11  12  16  25
    17  18  2  7  10
    4  23  20  3  1
    24  21  19  14  9
    6  22  8  13  15
```

У цьому прикладі краща (найменша) медіана рангів якості 9 досягається у середньому правому прямокутнику з елементів масиву Q , що виділені жирним шрифтом. Отже, `rectangle(R,C,H,W,Q)=9`.

Приклад 2

```
R=2, C=6, H=1, W=5,
Q=  6  1  2  11  7  5
    9  3  4  10  12  8
```

У цьому прикладі правильна відповідь — 5.

Підзадача 1 [20 балів]

R і C не перебільшують 30.

Підзадача 2 [20 балів]

R і C не перебільшують 100.

Підзадача 3 [20 балів]

R і C не перебільшують 300.

Підзадача 4 [20 балів]

R і C не перебільшують 1 000.

Підзадача 5 [20 балів]

R і C не перебільшують 3 000.

Рекомендації щодо розв'язання

Ця задача виглядає як одна з багатьох задач на решітці. Здається придатним складний алгоритм діапазонного пошуку (*range-search*), але існує значно простіший алгоритм, що розв'язує задачу на 100%.

Нехай $N=R \times C$ визначає розмір задачі.

Підзадача 1

Підзадачу 1 можна розв'язати найбільш очевидним алгоритмом, який розглядає кожен прямокутник (усього таких $(R-H+1) \times (C-W+1)$), сортує за квадратичний час їх вміст $((H \times W)^2$ кроків), знаходить медіанний ранг та обирає оптимальний. Найгірший випадок буде, коли $H=R/2$ та $W=C/2$. Отже, складність такого алгоритму $O(N^3)$.

Підзадача 2

Використання якогось з $O(N \log N)$ алгоритмів сортування замість квадратичного покращує оцінку складності до $O(N^2 \log N)$, чого достатньо для розв'язання задачі 2.

Також, можливо використовувати сортування за $O(N)$ (сортування підрахунком), щоб отримати простий $O(N^2)$ алгоритм, але цього не достатньо для розв'язання підзадачі 2.

```
const
  MaxDimension = 3000;
  MaxRank = sqrt(MaxDimension);
type
  TCoordinate = 0 .. MaxDimension-1;
  TRank = 1 .. MaxRank;
  Qtype = array [ TCoordinate, TCoordinate ] of TRank;
  TRankSet = array [ TRank ] of Boolean;
var
  sorttemp: TRankSet; { для сортування та пошуку
медіани,
зроблено глобальним для швидкості }
function rectangle(R, C, H, W: Longint; var Q: Qtype):
TRank;
{ повертає кращу медіану з усіх HxW
прямокутників з RxC міста Q }
function median(a, b: TCoordinate): TRank;
{ повертає гран медіани у прямокутнику
з верхнім-лівим кутом у [a, b] }
var
  i, j: TCoordinate; { проходить по Q }
  r: Longint; { проходить [ 0 ] + sorttemp }
  c: Longint; { пахує елементи у sorttemp <= r }
begin
  { вставляє елементи з прямокутника у пучки }
  for i := a to a + H-1 do begin
    for j := b to b + W-1 do begin
      sorttemp[ Q[i, j] ] := True
    end { for j }
  end { for i }
  { визначає медіану підрахунком половини елементів }
  r := 0 { r в [ 0 ] + sorttemp }
  for c := 1 to (H * W) div 2 + 1 do begin
    repeat r := r + 1 until sorttemp[r]
  end { for c }
  median := r
  { відбудовує інваріант для sorttemp,
вилучаючи елементи }
  for i := a to a + H-1 do begin
```

```

for j := b to b + W - 1 do begin
  sorttemp[Q[i, j]] := False
end { for j }
end { for i }
end; { median }
var
i: TRank; { проходить sorttemp для ініціалізації }
result: TRank; { поточний кращий знайдений ранг }
a, b: TCoordinate; { проходить всі прямокутники }
m: TRank; { медіана прямокутника з верхнім-лівим
кутом в [a, b] }
begin
for i := 1 to R * C do sorttemp[i] := False
; result := MaxRank;
; for a := 0 to R - H do begin
for b := 0 to C - W do begin
m := median(a, b)
; if m < result then result := m
end { for b }
end { for a }
; rectangle := result;
end; { rectangle }

```

Існують деякі очевидні можливості для покращення, такі як використання суттєвого перекриття між деякими прямокутниками під час заповнення або звільнення масиву перед сортуванням. Але ці покращення не впливають суттєво на оцінку складності.

Підзадача 3

Існує $O(N^{1.5} \log N)$ розв'язує підзадачу 3. Один із підходів такий. Нехай відомо вертикальне розташування шуканого прямокутника [$O(N^{0.5})$ можливостей]. Тоді будемо проходити по рядку, використовуючи якусь ефективну структуру даних, щоб слідувати за медіаною (можна використовувати дерево відрізків та двійковий пошук або пару куч для значень більше та менше). [Кожне значення додається та видаляється зі структури даних один раз, усього потрібен час $O(N \log N)$].

Цей підхід потребує великої кількості кодування.

Підзадача 4

Ця підзадача розрахована на можливі розв'язки складності $O(N \log N \log N)$.

Підзадача 5

Існує розв'язок складності $O(N \log N)$. Можна помітити, що результат роботи програми можна перевірити деяким алгоритмом, який відповідає на запитання «Чи є прямокутник, медіана якого $\leq X$?». Цей запит можна виконати за час $O(n^2)$. Прямокутник має медіану $\leq X$ тоді і тільки тоді, коли він містить більше значень $\leq X$ ніж навпаки. Присвоїмо всім клітинам решітки 'значення' відповідно до 'порогової' функції: -1 , якщо більше X , 0 , якщо дорівнює X , 1 , якщо менше X . Використовуючи відомі «накопичувальні» структури даних для запитів на прямокутних структурах, випробуємо всі можливі розташування прямокутника та будемо повертати «так», якщо 'значення' всередині якоїсь суми ≥ 0 . Простим бінарним пошуком по X знайдемо мінімальне значення, для якого відповіддю буде «так».

Такий $O(N \log N)$ алгоритм отримує 100% балів.

Лінійний Розв'язок

Твердження 1. Маючи значення X , можна за час $O(HW)$ встановити, які прямокутники мають медіану, кращу ніж X .

Доведення. За лінійний час побудуємо масив часткових сум: $A[i][j]$ = кількість елементів в $Q[0..i][0..j]$, кращих ніж X . Кількість елементів, кращих за X , у довільному прямокутнику тепер можна порівняти як комбінацію 4 значень з A . Медіана прямокутника розміром H на W краща X тоді і тільки тоді, якщо кількість елементів, кращих за X , менша ніж $HW/2$.

Твердження 2. Кращу медіану K прямокутників можна визначити за час $O(K^2 \log(HW) + HW)$.

Доведення. Зведемо все до сітки $K \times K$ та бінарним пошуком знайдемо кращу медіану. На кожному кроці потрібно пройти по всій сітці $K \times K$, а кількість елементів на початку HW , але зменшується геометрично на кожному кроці.

Твердження 3. Можна знайти кращу медіану всіх прямокутників за час $O(HW)$.

Доведення. Встановимо $K = \sqrt{HW} / \log(HW)$, щоб отримати лінійний час. Візьмемо K прямокутників, застосуємо *Твердження 2*. За *Твердженням 1* відфільтруємо всі прямокутники з гіршою медіаною. Залишиться HW/K прямокутників. Зробимо те ж саме ще раз, залишиться $HW/K^2 \leq \log(HW)$ прямокутників. Тепер знову застосуємо *Твердження 2*.

4. МОВИ

Вам необхідно написати інтерактивну програму, яка для даної послідовності цитат із Вікіпедії (див. приклад нижче) вгадує для кожної з них мову, на якій вона написана, у порядку їх черги. Після кожної здогадки вашої програмі дається правильна відповідь, яка може бути використана для покращення припущень вашої програми протягом її виконання.

Кожна мова задається цілим числом L від 0 до 55. Кожна цитата складається рівно із 100 символів і є масивом, елементами якого є 100 цілих чисел у межах від 1 до 65535. Ці цілі числа від 1 до 65535значаються довільно, і не відповідають ніякому стандартному кодуванню.

Вам необхідно реалізувати процедуру $\text{excerpt}(E)$, де E — це масив із 100 чисел, що є цитатою з Вікіпедії в описаному вище форматі. Кожен виклик процедури $\text{excerpt}(E)$ має викликати процедуру $\text{language}(L)$ рівно один раз, де L — припущення відносно мови статті з Вікіпедії, з якої було отримано дану цитату E . Система оцінювання реалізує процедуру $\text{language}(L)$, яка оцінює вашу здогадку і повертає правильну мову даної статті. Таким чином, ваше припущення є правильним, якщо $\text{language}(L) = L$.

Система оцінювання викликає процедуру $\text{excerpt}(E)$ 10 000 разів, один раз для кожної цитати з вхідного файлу. *Точність* вашого розв'язку буде визначатись як відношення P/Q , де P — кількість цитат, для яких реалізована вами процедура $\text{excerpt}(E)$ правильно визначила мову, Q — загальна кількість цитат, для яких викликала реалізована вами процедура $\text{excerpt}(E)$.

Ви можете використовувати різні методи для розв'язання цієї задачі. Метод Roschio — це один із способів, що дозволяє досягти точності біля 0.4. Метод Roschio визначає схожість цитати E для кожної з відомих мов L і обирає ту мову, яка найбільш схожа. Схожість визначається як загальна кількість різних символів із цитати E , що з'являються серед попередніх цитат на мові L .

Зверніть увагу, що вхідні дані були завантажені з реальних статей Вікіпедії, і тому в них можуть з'являтися некоректні символи або фрагменти тексту. Це є цілком очікуваним і є частиною цієї задачі.

Приклад. Як ілюстрації ми приводимо текстову презентацію цитат із 56 різномовних статей.

Приклад вхідного файлу grader.in.1 містить 10 000 таких цитат. Ці 56 мов — це мови, що були заявлені як рідні при реєстрації учасників IOI 2010. Мова кожної цитати обирається довільно із цих 56 мов, і кожна цитата обирається з першого параграфу статті, обраною випадково з Вікіпедії на відповідній мові. Кожен рядок файлу містить:

- дволітерний ISO-код відповідної мови Вікіпедії;
- 100 цілих чисел від 1 до 65 535, що подають цитату, складену з перших 100 символів з першого параграфу статті;
- додатне для перегляду представлення у кодуванні UTF-8 цих 100 символів, яке ви можете проглянути в ва-

шому текстовому редакторі, або у браузері Firefox (це представлення надається тільки для вашої зручності, і не призначено для використання у якості входу для вашої програми).

Система оцінювання журі використовує 10 000 різних цитат, обраних тим самим способом з Вікіпедії на тих самих 56 мовах. Однак, система оцінювання журі признає інші номери від 0 до 55 кожній мові й інші числа від 1 до 65 535 кожному символу.

Підзадача 1 [30 балів]

Ваш розв’язок має робити з точністю 0.3 або краще на офіційній системі оцінювання.

Підзадача 2 [до 80 балів]

Ваші бали будуть розраховані за формулою $114(\alpha - 0.3)$ і заокруглені до найближчого цілого. Тут α — це точність вашого розв’язку.

Рекомендації щодо розв’язання

Є багато підходів для розв’язання. Методу *Rocchio*, який описано в умові задачі, достатньо для розв’язання підзадачі 1.

Для підзадачі 2 не достатньо просто підраховувати, з якою частотою зустрічаються символи. Необхідно збирати статистику для біграм (пари сусідніх символів), триграм (три послідовних символи), що призведе до більшої охайності.

Таблиця 1

0.	Yshokkie word meestal in Kanada , die noorde van die VSA en in Europa gespeel. Dit is bekend as 'n b
1.	وهو المنتج الذي يجعل المنظم لا يكسب ربحا ولا يخسر ويحصل على دخل يكف ، Marginal Producer المنتج الحدي
2.	"BAKIL" Futbol Klubu 1995-ci ildə Misir Səttar oğlu Əbilov tərəfindən yaradılmış və həvəskar futbol
3.	Квинт Фулвий Флак (Quintus Fulvius Flaccus; † 205 пр.н.е.) е политик и генерал на Римската република
4.	ইন্ডিয়ান ইম্পিট্রিট অফ গেশাল ওয়েলফেয়ার আন্ড বিজনেস ম্যানজমেন্ট (সংক্ষেপে আইআইএসমডিইটিএম)
5.	5. juni (lipanj) (5.6.) je 156. dan godine po gregorijanskom kalendaru (157. u prestupnoj godini
6.	La Caunette és un municipi francès , situat al departament de l' Erau i a la regió de Llenguadoc-Ros
7.	Praha je malé městečko v Texasu , které leží cca 85 km na jihozápad od Austinu . Bylo založeno
8.	Graeme Allen Brown (født 9. april 1979 i Darwin , Northern Territory , Australien) er en australsk
9.	Der Plattiger Habach (3.214 m ü. A . , nach anderen Angaben nur 3.207 m [1])
10.	To Νησί Γ κρέη Μπάριερ (Αγγλικά : Great Barrier Island , Μαορί : Motu Aotea) είναι νησί στα βόρεια
11.	Sid Bernstein Presents... is a 2010 feature-length documentary film by directors Jason Ressler and E
12.	El término latino lex loci celebrationis aplicado al derecho internacional privado quiere decir: "le
13.	Apollo 5 oli kosmoselaev , mis sooritas Apollo programmi teise mehitamata lennu. Lennu käigus testit
14.	1350 : هزار و سیصد و پنجاهمین سیارک کشف شداس (آمارگناری : 1934 : Rosselia سیارک ، ۱۳۵۰ (به انگلیسی :
15.	V. I. Beretti (mýös Vikenty Ivanovitš Beretti , alk. Vincent Beretti ; 1781 Milano Italia – 18. elok
16.	Le 5 e bataillon de parachutistes vietnamiens (ou 5 e BPVN ou encore 5 e Bawouan) est une unité par
17.	Amina Sarauniyar Zazzau , wadda ta rayu daga shekarar 1533 zuwa 1610 , daya ce daga cikin 'ya'ya biyu
18.	ב מהמטביקה , השערת רימן היא השערה שיש צפיפות של מספרים ראשוניים הולכת לאפס של האינסוף
19.	Sudski proces Doe protiv Boltona je sudski proces iz 1973 . godine kojim je američki Vrhovni sud uki
20.	Owen Cunningham Wilson (1968 . november 18 . , Dallas , Texas , Egyesült Államok) amerikai színész
21.	Հայ Վարդանիկե Եկեղեցին պատկանում է Արևելյան Վարդանիկ Եկեղեցիներին և այսպիսով ենթարկվում է Հռոմի Պապի ա
22.	Dionysios dari Halicarnassus (Bahasa Yunani : Διονύσιος Αλεξανδρού Αλικαρνασσεύς , Dionysios putra
23.	Nnamdi "Zik" Azikiwe , bu onye isi-ala izizi Nijiria nwere. Ochihi na afọ 1954 welu ruo n
24.	La Riserva naturale orientata Serre della Pizzuta è un'area protetta del dipartimento Regionale di S
25.	石橋和義 (いしばし かずよし/まさよし、生没年不詳) は、 □□詳。石橋氏 初代当主。初名氏義。尾張 三郎を 通称とし、官途は、左近将監 → 三河守 → 左衛門佐。足利直義
26.	კორბინ ბლეუ (ინგლ. Corbin Bleu ; დ. 21 თებერვალი , 1989 , დაბადების ადგილი ბრუკლინი , ნიუ-იორკი , ა
27.	Тарья Каарина Халонен (Tarja Kaarina Halonen); 24 желтоқсан , 1943 , Каплио , Хельсинки , Финлан
28.	딜롱 (Dilong)은 중국 라오닝(Liaoning) 지방의 익시안층(Yixian Formation)에서 온전한 4구의 화석으로 발견되었 다. 이 공룡은 가장 원시적인 초기의 티
29.	Сүймөнкул Чокморов - советтик актёр. Жетинин айынын 9 (ноябрь) 1939-жылы, Фрунзе шаарын жанындагы
30.	D' Mirjam vun Abellin war eng Nonn a Mystikerin , och " Maria vum gekräizegte Jesus " genannt. Si as
31.	Panorea abrupta (angl. Geoduck) - jūriņiņ dvgeldziņ moliuskų rūšis, priklausanti Hiatellidae šeim
32.	"Dzimis Latvijā" ir Liepājas dueta Fomins & Kleins 2004 . gada 23. februārī izdotais otrais albu
33.	I Ludwik Lejzer Zamenhof dia dokotera mpijery maso nipetraka any Polonia . Fantantsika izy ankehitr
34.	Седумстотини милиони малечки алвеоли во белите дробови , всушност се шупливи чаури - алвеоли прекрие
35.	Энэхүү шувуу нь Бутан , Хятад , Гонконг , Энэтхэг , Пакистан , Иран , Япон , Казакстан , Солонгос ,
36.	भारतातील महाराष्ट्र राज्याच्या नागपूर पासून २१६ कि.मी. दूर असलेले एक गाव. ते वैनगंगा नदीच्या काठावर
37.	De Slotervaart was oorspronkelijk de waterweg die sinds de Mitteleeuwen het dorp Sloten verbond met
38.	Macierz S (macierz rozpraszania , od ang. scattering matrix) jest centralnym elementem w mechanic k
39.	A Hora do Rush 3 (Rush Hour 3 , no original) é o terceiro filme da franquia Rush Hour . Dirigido po
40.	Coorodate : 51°34'0"N 12°3'0"E / 51.56667 , 12.05 Brachstedt este o comună din landul Saxonia-A
41.	Гробници императоров династии Мин и Цин — памятник Всемирного наследия ЮНЕСКО , состоящий из несколько
42.	Kovalentni radijus atoma - ponekad se naziva i valentni radijus. Kovalentni radijus je srednje rasto
43.	Koniecpol je mesto v Polsku v Sliezskom vojvodstve v okrese Powiat częstochowski v rovnomennej gmine
44.	Hoxhë Vokrii vije nga Shqipëria ishte një klerik shqiptar i cili luftonte për Çështjen Kombëtare . A
45.	Гурдијелъе је насеље у општини Тутин у Рашком округу . Према попису из 2002. било је 93 становника (п
46.	Underhållsstöd betalas ut av Försäkringskassan (FK) till en förälder som är vårdnadshavare och bor e
47.	இந்தியாவின் தேசிய நெடுஞ்சாலைகள் நடுவண் அரசின் தேசிய நெடுஞ்சாலைத் துறையால் பராமரிக்கப்படுகின்றன. பெரு
48.	Дар он зиндаги . машат , фаолияти мехнати,муборизаи ичтимои, русуму омом, хислат ва эҳсосоти хақк ифто
49.	ໄພທອພອວາ ອິນເຟສທັນສ (อังกฤษ : Phytophthora infestans) คือเชื้อ ราโอไมซ์ที่ หรือ ราน้ำ ที่เป็นสาเหตุ
50.	ABUL FAWARIS BERRANY - 11. asyrda Orta Aziyadaky oguz taýpalarynyň berrany dinastiasynyň wekili. Ol
51.	Egemenlik ya da hâkimiyet , bir toprak parçası ya da mekan üzerindeki kural koyma gücü ve hukuk yara
52.	Темне фентези (від англ. Dark Fantasy - темне, похмуре фентези) - піджанр літератури, який включає
53.	Paris By Night 84: In Atlanta - Passport to Music & Fashion (Am nhạc và Thời trang) là chương tr
54.	ISO 3166-2:GU ni akooole ninu ISO 3166-2 , apa opagun ISO 3166 ti International Organization for Stan
55.	下卡姆斯克 (俄文 : Нижнекамск ; 鞑靼語 : Түбән Кама/Tübän Kama) 是 俄羅斯 鞑靼斯坦共和國 東北部的 一個城市 , 位於 卡馬河 南岸。2002年 人口22

(Далі буде)