Vinjetter
TDDE22 Datastrukturer och algoritmer

Magnus Nielsen
Filip Strömbäck
Tommy Färnqvist

25 augusti 2023

Scenario 1

Man har inom Posten Logistik AB skrivit programvara för sortering av kundinformation och vill standardisera användningen av sorteringsalgoritmer. Som ett första steg har man provkört de olika algoritmerna på samma dator. Man har provat data av olika storlek och uppnått följande resultat:

storlek indata	alg. 1	alg. 2	alg. 3	alg. 4
10	0.00044	0.00057	0.00041	0.00046
100	0.00675	0.00420	0.00171	0.00244
1000	0.59564	0.05565	0.02927	0.02587
10000	58.864	0.71650	0.42998	0.31532
100000	∞	8.8591	5.7298	3.5882
1000000	∞	104.68	71.164	41.282

Scenario 2

ADT Map/Dictionary har många tillämpningar. Följande är några exempel:

1. Uppslagstabell för att beräkna $\sin(\theta)$, där θ är en av 1000000 möjliga vinklar jämnt fördelade mellan 0 och π.
2. Databas som avbildar ljuddata (från en fil) på artistnamn.
3. Snabbaste garantierna för insättning, borttagning och sökning för en godtycklig uppsättning numeriskt data.

Spelar det någon roll vilken underliggande implementation man väljer i ovanstående tllämpningar?

Scenario 3

En rekursiv implementation av *fibonacci*-funktionen är dyr att exekevera. Den kan implementeras enligt nedan:

$$\text{fib}(x) = \begin{cases} 0 & \text{om } x = 0 \\ 1 & \text{om } x = 1 \\ \text{fib}(x-1) + \text{fib}(x-2) & \text{annars} \end{cases}$$

Det finns flera tekniker för att beräkna *fibonacci* mer effektivt. Ett alternativ är att använda *dynamisk programmering* (*memoisation*). Ett annat alternativ är att utnyttja det faktum att *fibonacci* är en linjär funktion (dvs. en *linear recurrence*). Den går alltså att uttrycka enligt nedan (T är en matris, x är vektorer):

$$x_n = Tx_{n-1}$$

Dessa implementationer har olika karaktäristik och går därmed att applicera i olika situationer.

Tips: Matrisexponentiering (T^n) kan implementeras effektivt med insikten att för jämna n så gäller: $T^n = T^{n/2} * T^{n/2}$. Vi kan alltså implementera exponentiering för heltalsexponenter effektivt endast med hjälp av matrismultiplikation.

Scenario 4

Du är på besök i ett land du aldrig tidigare besökt. Till din hjälp har du fått en lista på alla matställen i staden. Listan innehåller n (mycket stort) rader, varje rad innehåller namn på restaurang, namn på en rätt samt pris på rätten. Du kan inte språket, men du har lyckats lista ut att ett V efter namnet inntebär vegetariskt, G innebär glutenfritt.

Exempel på lista:

Restaurang	**Namn på rätten**	**Pris**
Dorsia	Seekoeier ceviche	17
Kolm luud	Pivo od maslaca V G	3
Los Pollos Hermanas	Pollo Caliente G	15
Kafejo Monk	Velika Salata V G	19
Los Pollos Hermanas	Pollo Frio G	10
...	...	...

Scenario 5

Dagbrott AB vill bryta malm för att maximera sin nettovinst. Området där dagbrottet ska anläggas är indelat i block och för enkelhets skull antar vi en endimensionell layout. Från den geologiska prospekteringen är bedömningen att brytning av block `i` ger nettovinst (värdet av malmen minus utvinningskostnaderna) a_i miljoner kronor. Vissa miljörelaterade restriktioner gör att Dagbrott AB endast får bryta malm i en sammanhängande grupp av block. Vilka block bör Dagbrott AB välja? Med andra ord, givet en sekvens av `N` (möjligen negativa) heltal, bestäm den största möjliga summan bland alla sammanhängande sekvenser. Problemet är lätt om alla tal är positiva: välj hela sekvensen. Svårigheten är när det förekommer negativa tal: borde man ta med ett negativt tal i hopp om att närliggande positiva tal kompenserar för detta?

För indatat nedan är den maximalt möjliga summan 104, vilken erhålls genom att ta med block 2 till 6. Notera att summan alltid är minst 0, eftersom Dagbrott AB kan välja att inte gräva alls.

block	0	1	2	3	4	5	6	7	8	9	10
nettovinst	12	-34	40	6	-10	56	12	-1	-15	10	4

Tillämpningar

I verkliga tillämpningar behöver gruvoperatörer lösa en tredimensionell variant av problemet. Dessutom kan det förekomma ytterligare geologiska begränsningar, t.ex. att det inte går att bryta ett område 100 meter under marken utan att först frilägga några av de omgivande områdena. Den tvådimensionella varianten av problemet förekommer även som del av bildbehandling, där målet är att bestämma maximum likelihood-skattningen för ett visst sorts mönster.

Exempeldata

På katalogen /courses/TDDE22/vinjetter/pitmining/ finns flera filer med exempeldata.