NTNU
Kunnskap for en bedre verden

Institutt for datateknikk og informasjonsvitenskap

Eksamensoppgave i TDT4120 Algoritmer og datastrukturer

Faglig kontakt under eksamen	Magnus Lie Hetland
Telefon	918 51 949
Eksamensdato	9. august, 2017
Eksamenstid (fra–til)	09:00–13:00
Hjelpemiddelkode/tillatte hjelpemidler	D
Annen informasjon	Oppgavearkene leveres inn, med svar i svarrute under hver oppgave
Målform/språk	Bokmål
Antall sider (uten forside)	5
Antall sider vedlegg	0

Informasjon om trykking av eksamensoppgave

Originalen er

1-sidig	☑	**2-sidig**	☐
sort/hvit	☑	**i farger**	☐

Skal ha flervalgskjema ☐

Kvalitetssikret av Pål Sætrom

Kontrollert av

Dato Sign

Merk: Studenter finner sensur i Studentweb. Har du spørsmål om din sensur må du kontakte instituttet ditt. Eksamenskontoret vil ikke kunne svare på slike spørsmål.

Les dette nøye

(i) Les hele eksamenssettet nøye *før* du begynner!

(ii) Faglærer går normalt én runde gjennom lokalet. Ha evt. spørsmål klare!

(iii) Skriv svarene dine i svarrutene og levér inn oppgavearket. Bruk gjerne blyant! Evt. kladd på eget ark først for å unngå overstrykninger, og for å få en egen kopi.

(iv) Ekstra ark kan legges ved om nødvendig, men det er meningen at svarene skal få plass i rutene på oppgavearkene. Lange svar teller ikke positivt.

(v) Eksamen har 15 oppgaver, totalt verdt 100 poeng. Poengverdi er angitt ved hver oppgave.

Merk: Varianter av de foreslåtte svarene nedenfor vil naturligvis også kunne gis uttelling, i den grad de er helt eller delvis korrekte.

Oppgaver

(5 p) 1. Hva er *worst-case*-kjøretiden til Insertion-Sort? Oppgi svaret i Θ-notasjon.

$\Theta(n^2)$

(5 p) 2. Din venn Lurvik har et program som sorterer lenkede lister, med en tilpasset versjon av Insertion-Sort. Hun har nettopp lært seg Merge-Sort, og har lyst til å bruke den, men er bekymret for at den ikke vil være så effektiv om den tilpasses lenkede lister. Hva mener du? Forklar.

Med lenkede lister vil det fortsatt ta lineær tid i tillegg til de to rekursive kallene, så kjøretiden blir fortsatt $\Theta(n \lg n)$, og den er altså mer effektiv enn Insertion-Sort.

Her får man noe uttelling for å resonnere rundt forskjeller på arrays og lenkede lister, og kostnad ved direkte oppslag på midten i Merge-Sort, etc.

(5 p) 3. Du har oppgitt en tabell $A = \langle 1, 7, 2, 3, 5, 4, 6 \rangle$.

Utfør Build-Max-Heap(A) og deretter Heap-ExtractMax(A). Oppgi A etterpå.

Merk: I svaret så er $A.heap\text{-}size = A.length - 1$. Det spørres her om hele A, ikke bare heapen.

$\langle 6, 5, 4, 3, 1, 2, 7 \rangle$

(5 p) 4. Hvilke to egenskaper ved et problem ser vi etter for å avgjøre om vi vil bruke dynamisk programmering (DP)? Oppgi først egenskapen som er nødvendig for at DP skal være *mer effektivt* enn naturlige alternativer, og deretter egenskapen som er nødvendig for at DP skal gi *korrekt svar*.

Overlappende delproblemer og optimal substruktur/delstruktur.

Noe uttelling gis også for å oppgi begge, men i feil rekkefølge.

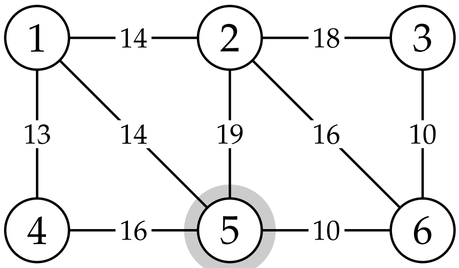
(5 p) 5. Du skal konstruere en Huffman-kode for tegnene a–d, med frekvenser som angitt nedenfor.

Tegn	a	b	c	d
Frekvens	4	1	6	2

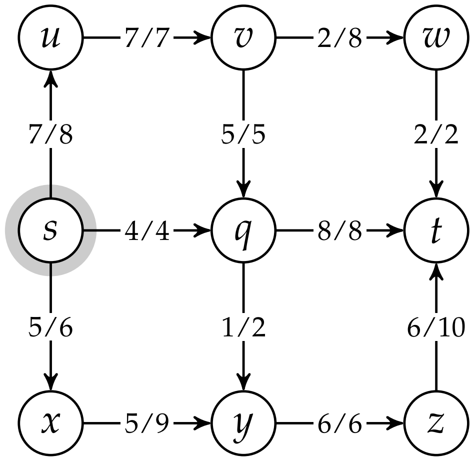
Hva blir kodeordet for b?

Merk: Venstre barn plukkes ut først, og kanten til venstre barn får verdien 0.

100

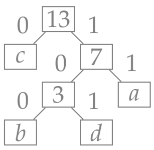


Figur 1: En vektet, urettet graf, brukt i oppgave 6



Figur 2: Flytnettverk brukt i oppgave 9

Først plukkes b og d ut, og får samlet vekt 3. Deretter plukkes (b, d) og a ut, og får vekt 7. Deretter plukkes c og $((b, d), a)$ ut, og får vekt 13. Vi får da følgende tre:



(5 p) 6. Hvis du utfører MST-PRIM på grafen i figur 1, med 5 som rot, hvilken kant vil velges som den femte i rekken? Det vil si, hvilken kant vil være den femte som legges til i løsningen?

Oppgi kanten på formen (i, j), der $i < j$.

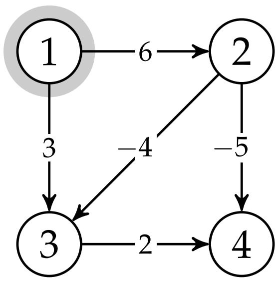
$(1, 2)$

(5 p) 7. Din venn Smartnes vil finne korteste vei fra node 1 til node 4 i grafen i figur 3, og har bestemt seg for å bruke DIJKSTRA, med 1 som startnode. Oppgi avstandene han finner til nodene 2, 3 og 4, i rekkefølge.

Her godtas to svar: 6, 3, 5 eller 6, 2, 1.

Her finner DIJKSTRA riktig svar, tross negative kanter, men det er bare «flaks». Hadde f.eks. $w(2, 4)$ vært høyere, ville 4 ha fått gal verdi. Normalt har DIJKSTRA funnet avstanden til en node idet den plukkes ut av Q, og tanken var egentlig at man skulle oppgi disse (6, 3, 1), som altså er gale. (de blir forbedret senere når vi besøker 2.) Men siden dette ikke er éntydig spesifisert, så godtas også verdiene d etter at algoritmen har kjørt ferdig, som egentlig er et mer naturlig svar, slik oppgaven er formulert.

Om man kun har svart at Dijkstras algoritme ikke fungerer på negative tall, så gis det 1 poeng.



Figur 3: En vektet, rettet graf til bruk i oppgave 7

$D^{(3)}$

	1	2	3	4
1	0	−3	5	−1
2	∞	0	∞	2
3	∞	−8	0	−6
4	∞	−5	3	−3

$\Pi^{(3)}$

	1	2	3	4
1	NIL	3	1	2
2	NIL	NIL	NIL	2
3	NIL	3	NIL	2
4	NIL	3	4	2

Figur 4: Forrige tilstand i utførelsen av Floyd-Warshall, brukt i oppgave 8

(5 p) 8. Lurvik og Smartnes har slått seg sammen, for å finne de korteste veiene mellom *alle* noder i en rettet, vektet graf. De har valgt å bruke Floyd-Warshall, men etter å ha utført den manuelt i et par iterasjoner har de begynt å lure på om det er noe galt med grafen deres. Du har sagt du skal hjelpe dem med neste iterasjon.

$D^{(3)}$ og $\Pi^{(3)}$ er som angitt i figur 4. Hva blir $D^{(4)}$ og $\Pi^{(4)}$?

Fyll ut tabellene nedenfor.

Merk: Vi antar her en implementasjon som i læreboka, det vil si at vi i hver iterasjon k lager nye tabeller $D^{(k)}$ og $\Pi^{(k)}$, heller enn en mer plass-effektiv variant som overskriver tabellene.

Her er det noe galt med grafen, som Lurvik og Smartnes mistenkte: Den inneholder en negativ sykel. Det betyr ikke at det ikke er mulig å utføre algoritmen, naturligvis – eller det neste trinnet i algoritmen, som dere blir bedt om.

Merk: Svartabellene var også merket med $D^{(3)}$ og $\Pi^{(3)}$ i eksamenssettet (altså en skrivefeil).

Selv om dette ikke bør ha påvirket svaret man gir, så kan det ha ført til forvirring og tap av tid, og derfor tas oppgaven ut av sensur der det fører til økt poengsum.

$D^{(4)}$

	1	2	3	4
1	0	−6	2	−4
2	∞	−3	5	−1
3	∞	−11	−3	−9
4	∞	−8	0	−6

$\Pi^{(4)}$

	1	2	3	4
1	NIL	3	4	2
2	NIL	3	4	2
3	NIL	3	4	2
4	NIL	3	4	2

(5 p) 9. Figur 2 på side 2 viser flytnettverket G, med kilde s, sluk t og flyt f. Er flyten maksimal? Svar ja eller nei.

Hvis ja, oppgi også mengden av noder som kan nås (dvs., som det finnes stier til) fra s i G_f.

Hvis nei, oppgi også nodene i en flytforøkende sti (*augmenting path*), i rekkefølge.

Ja: u, v, w, s, q, x, y

Her får man også full uttelling om man ikke har tatt med s. Man får også noe uttelling om man ikke har fått med seg flytopphevingen, og har droppet v og q (men mindre uttelling dersom man har droppet bare én av dem). Om man har tatt med t, vitner det om fullstendig manglende forståelse for det oppgaven spør om, og det gis ingen uttelling.

(5 p) 10. Stemmer det at $\mathrm{P} \subseteq$ co-NP? Forklar svært kort.

Ja: Om et språk kan avgjøres i polynomisk tid, så kan også komplementet avgjøres, og dermed verifiseres, i polynomisk tid.

(10 p) 11. Hvilke sorteringsalgoritmer i pensum har lineær forventet (*average-case*) kjøretid, under normale antagelser? Hva er disse antagelsene? (Det holder med et par stikkord per algoritme.)

COUNTING-SORT:	Input er heltall $0 \dots k$.
RADIX-SORT:	Input har d siffer $0 \dots k$
BUCKET-SORT:	Input er uniformt fordelte over $[0, 1)$

(10 p) 12. En *uavhengig mengde* (*independent set*) i en graf $G = (V, E)$ er en delmengde $U \subseteq V$ av nodene som er slik at hver kant i E er tilkoblet maksimalt én node i U (dvs., ingen av nodene i U er naboer). *Størrelsen* til en uavhengig mengde er antall noder den inneholder. Vis at følgende problem er NP-komplett.

$$\text{INDEPENDENT-SET} = \{\langle G, k\rangle : G \text{ er en graf med en uavhengig mengde med størrelse } k\}$$

Det holder med en kort forklaring, men alle elementene i et NP-kompletthetsbevis må dekkes.

Selve beviset er en svært enkel reduksjon fra CLIQUE: Komplementér kantmengden. Det viktigste er at man beskriver en reduksjon *fra* og ikke *til* CLIQUE. I tillegg bør man kort si noe om hvorfor INDEPENDENT-SET er i NP, og at reduksjone gir en ekvivalens (*hvis og bare hvis*).

Om man har med bare noen av elementene, vil det kunne gi noe uttelling. Om man reduserer feil vei, teller det svært negativt.

(10 p) 13. Du har oppgitt et sett med regler av følgende type, som beskriver en ukjent mengde S, der S er en delmengde av $\{1, \dots, n\}$, for en gitt n:

«Hvis x_1 eller x_2 eller $\cdots$ eller x_{m-1} ligger i S så ligger x_m i S.»

Her er $x_1, \dots, x_m$ elementer i $\{1, \dots, n\}$, og m kan variere fra regel til regel. Anta at du får tid til å bygge en datastruktur basert på reglene. Deretter skal du effektivt kunne løse følgende problem:

Input: En verdi x som skal ligge i S.
Output: Hele mengden S.

Her skal S være den *minste mengden* som tilfredsstiller opplysningene du har fått. Det vil si, S inneholder x og akkurat de elementene som kreves av reglene, men ingen andre. Du finner et eksempel på neste side.

Beskriv hvordan du vil løse dette problemet så effektivt som mulig.

Bygg en rettet graf, der hver regel angir kantene inn til node x_m. Traversér fra oppgitt element.

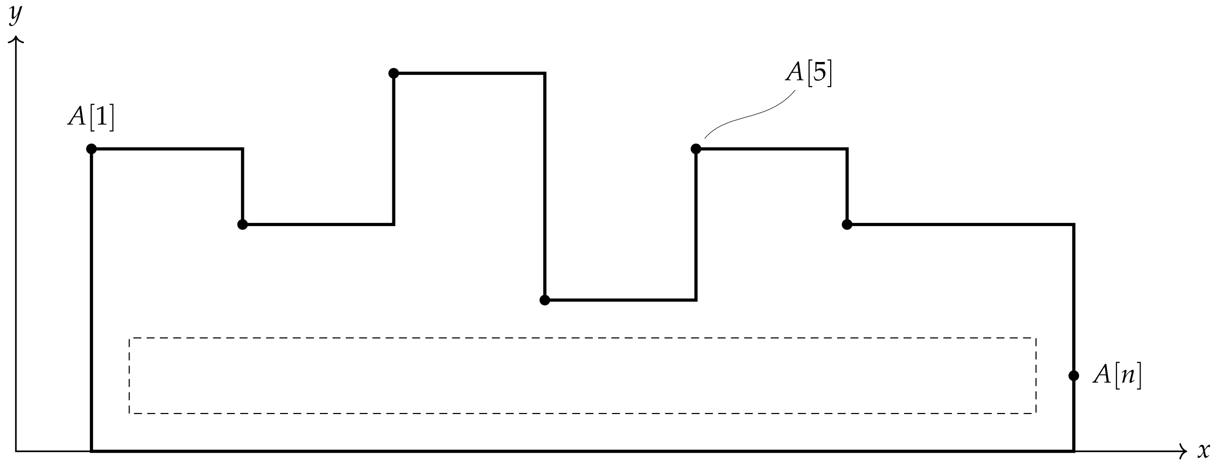
Om oppgitt element ikke er en node i grafen, så vil S bare inneholde det oppgitte elementet. Det er ikke kritisk at man påpeker dette.

Man kan godt for eksempel kjøre TRANSITIVE-CLOSURE på grafen først, men det vil ikke endre den asymptotise kjøretiden.

(10 p) 14. Du har oppgitt en tabell $A[1 \,..\, n]$, der hvert element $A[i]$ er et punkt med positive koordinater (x_i, y_i) i planet, med $x_i < x_{i+1}$ for $i = 1 \dots n-1$. Du skal finne et rektangel med horisontale/vertikale sider som ligger innenfor regionen definert av punktene, som vist i figur 5 på neste side. Mer spesifikt så skal du finne det *største* rektanglet av denne typen, altså det med størst areal.

Beskriv en algoritme som løser problemet så effektivt som mulig. Det holder med en overordnet, stikkordspreget forklaring, uten grundige implementasjonsdetaljer. For enkelhets skyld trenger du bare finne *arealet* til løsningen, ikke koordinatene. Hva blir kjøretiden?

Splitt og hersk. Finn laveste punkt og gang med bredden. Løs deretter problemet rekursivt til venstre og høyre for laveste segment. Velg den beste av de tre. Med lineær kjøretid for å finne minimum blir kjøretiden i beste/forventede tilfelle $\Theta(n \lg n)$ og i verste tilfelle $\Theta(n^2)$.



Figur 5: Eksempel på punkter, brukt i oppgave 14. Den tykke streken angir regionen, og består av en horisontal strek etterfulgt av en vertikal strek fra hvert punkt $A[i]$ til $A[i+1]$, for $i = 1 \ldots n-1$; fra $A[1]$ og $A[n]$ går det linjer ned til x-aksen, som utgjør den nederste siden av regionen. Det stiplede rektanglet er et eksempel på en gyldig, men ikke optimal løsning.

(10 p) 15. Du har oppgitt en todimensjonal tabell $A[1\,..\,n, 1\,..\,n]$ med reelle tall. Du skal velge ut ett tall fra hver rad slik at (1) ingen av de utvalgte tallene er i samme kolonne som det utvalgte tallet i neste rad, og (2) summen av tallene er størst mulig. (For $A[i,j]$ er i raden og j kolonnen.) Beskriv en algoritme som løser problemet så effektivt som mulig. Hva blir kjøretiden?

Dynamisk programmering. Delproblem: Beste løsning som slutter med $A[i,j]$. For hvert delproblem, velg beste løsning fra andre kolonner i raden over.

Eksempel til oppgave 13. Anta at du får oppgitt følgende regler:

(i) Hvis 1 eller 3 ligger i S så ligger 2 i S.

(ii) Hvis 2 ligger i S så ligger 3 i S.

(iii) Hvis 1 eller 3 ligger i S så ligger 4 i S.

Du får oppgitt et element x som ligger i S, og skal liste opp elementene i S. Her er svarene for ulike x:

$$
\begin{aligned}
x = 1 &\implies S = \{1,2,3,4\} \\
x = 2 &\implies S = \{2,3,4\} \\
x = 3 &\implies S = \{2,3,4\} \\
x = 4 &\implies S = \{4\} \\
x = 5 &\implies S = \{5\}
\end{aligned}
$$

I hvert tilfelle inneholder altså mengden S det oppgitte elementet x, men også de elementene den *må* inneholde for at reglene (i)–(iii) skal gjelde. Merk at S ikke inneholder unødvendige elementer.