

DM02 Eksamen – Obligatorisk Opgave

Evolutionstræer

Anden del: Letteste udspændende træ

1 Problemet

I denne del-opgave skal I konstruere et evolutionstræ. Input er n strenge, som karakteriserer n forskellige dyrearter. Det kan f.eks. være gensekvenser, dvs. strenge bestående af bogstaverne A, C, G og T. Vi er interesserede i at lave et "stamtræ" eller *evolutionstræ* for de n dyrearter. Ideen er, at hvis editingsafstanden mellem to strenge er lille, så er der en god sandsynlighed for, at de to dyrearter er i tæt familie med hinanden.

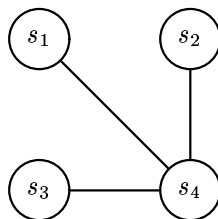
Der findes mange algoritmer til at gætte på, hvordan det virkelige evolutionstræ ser ud. En af de simpleste går ud på at betragte grafen med n knuder, som hver repræsenterer en af de n strenge. Hvert par af knuder er forbundet af en kant, hvis vægt er lig editingsafstanden mellem de to tilsvarende strenge. I denne graf findes et letteste udspændende træ. Da den samlede vægt af kanterne i et sådant træ er mindst muligt, håber man på, at træet fortæller noget om dyrearternes indbyrdes slægtskab. Det er sådan et træ, I skal konstruere.

1.1 Et lille eksempel

Betragt de fire strenge $s_1 = \text{AGATA}$, $s_2 = \text{TAG}$, $s_3 = \text{CAT}$ og $s_4 = \text{AGT}$. Deres indbyrdes editingsafstande er

	s_1	s_2	s_3	s_4
s_1	0	4	3	2
s_2	4	0	2	2
s_3	3	2	0	2
s_4	2	2	2	0

Et letteste udspændende træ har vægt 6 og kunne f.eks. se sådan ud:



2 Programmet

Programmet skal bygge videre på det program, I lavede i første delopgave; dvs. den del af programmet, som beregner afstandene mellem strengene skal have køretid $O(n^2 \ell^2)$ og pladsforbrug $O(n\ell)$, hvor ℓ er antallet af tegn i den længste af input-strengene. Husk, at i denne del-opgave er n antallet af strenge.

I skal konstruere et letteste udspændende træ vha. Kruskals algoritme. Denne del af programmet skal have køretid $O(n^2 \log n)$ og pladsforbrug $O(n^2)$. Til at sortere kanterne er det i orden at bruge en af JAVAs indbyggede sorteringsmetoder.

2.1 Tekniske krav

Input: De n strenge, der skal sammenlignes, lægges i en fil med n linier; en streng per linie. Input-filen svarende til eksemplet i afsnit 1.1 ser altså sådan ud:

```
AGATA
TAG
CAT
AGT
```

Output: Programmet skal skrive et enkelt tal, som angiver den samlede vægt af et letteste udspændende træ. For eksemplet i afsnit 1.1 vil output være tallet 6. Bemærk, at dit program ikke må skrive andet end dette tal. Hvis programmet f.eks. skriver “Den samlede vægt er 6”, så fejler det, når vi kører den automatiske test.

Hovedklassen skal hedde EvolutionaryTree (stort “E” og “T”), og man skal kunne afvikle programmet på følgende måde:

```
java EvolutionaryTree <strengfil>
```

hvor filen `strengfil` indeholder de n strenge som beskrevet ovenfor.

Vi har ikke tid til at teste hvert enkelt program manuelt. Det er derfor et ufravigeligt krav, at jeres program lever op til specifikationerne i dette afsnit — check dette vha. `DM02check`.

3 Test

I må gerne antage, at input har det rigtige format, men det skal testes grundigt, at programmet virker korrekt på korrekt input.

Hvis I gerne vil have en ide om, hvad vi senere vil sige til det output, I producerer, kan I skrive `DM02check`. I skal, før I afgiver kommandoen, placere jer i det katalog, som alle jeres oversatte JAVA-programmer ligger i. Så vil jeres program blive kørt på samtlige testfiler i

```
/home/IMADA/courses/dm02/Tests
```

Hvis der findes fejl, bliver dette skrevet på standard output, ellers skrives der intet. Hvis der ikke findes fejl, betyder det ikke nødvendigvis, at jeres program fungerer helt korrekt, da testen ikke er udtømmende. At udføre `DM02check` uden fejl er altså *ikke* tilstrækkeligt til at konkludere, at programmet virker.

Programmet `DM02check` kan afbrydes med `Ctrl-c`.

4 Krav til rapport og program

Først og fremmest skal alle krav beskrevet i denne opgaveformulering naturligvis tilfredsstilles.

Programmet skal være velstruktureret og kommenteret i passende omfang. Rapporten skal indeholde en udskrift af hele programmet. Denne udskrift skal være identisk med det elektronisk afleverede program.

Rapporten skal indeholde en beskrivelse af de væsentligste valg, der er truffet i forbindelse med design og implementation, samt begrundelser herfor. Man skal i hovedtræk kunne forstå, hvordan programmet løser problemet, uden at læse koden.

Der skal argumenteres for, at programmet har den krævede køretid og pladsforbrug.

Det skal forklares, hvordan programmet er afprøvet; test-eksemplerne skal beskrives, og man skal forklare, ud fra hvilke kriterier de er valgt. Det er ikke tilstrækkeligt blot at have kørt `DM02check`.

Eventuelle mangler i program eller rapport skal beskrives.

Endelig skal rapporten underskrives.

5 Aflevering

Der skal afleveres en rapport på sekretariatet, og et program skal afleveres elektronisk, begge dele inden mandag d. 1. december 2003 kl. 10:00. Det er jeres eget ansvar i god tid at checke, at det er teknisk muligt at foretage den elektroniske aflevering, dvs. at I har et login på IMADA osv.

Den elektroniske aflevering foregår på følgende måde. Opret et katalog (directory), som indeholder alle jeres JAVA-filer til opgaven og intet andet. Stil jer i dette katalog, og afgiv kommandoen `aflever DM02`. Dette skal ske fra maskinen `tosca`. Dvs. gør følgende.

```
ssh toasca (afgiv password, hvis der bliver bedt om det)
cd <opgave-katalog>
aflever DM02
```

hvor `opgave-katalog` indeholder alle jeres JAVA-filer til opgaven og intet andet.

Bemærk, at I (inden afleveringsfristen) kan aflevere flere gange, hvis I fortryder en aflevering. Kun den sidste aflevering tæller (de andre slettes).

Rapporten skal afleveres på sekretariatet på Institut for Matematik og Datalogi. Denne opgaveformulering er vedhæftet en forside, som skal anvendes ved afleveringen. Husk for jeres egen skyld at få en kvittering (også vedhæftet) på, at I har afleveret.

Husk, at det er et krav, at det elektronisk afleverede program er præcis det samme program, der afleveres en udskrift af i rapporten.

6 Evaluering

For at kunne gå til eksamen i DM02 til januar skal man have bestået den obligatoriske opgave. Har man tidligere lavet en obligatorisk DM02-opgave og fået den godkendt, er det tilstrækkeligt.

De rettede opgaver bliver leveret tilbage ved øvelserne onsdag d. 10. december. De, der ikke har øvelser om onsdagen, kan hente deres opgave på sekretariatet fra kl. 14 d. 10. december.

Som udgangspunkt er der kun to mulige bedømmelser af opgaven: "godkendt" eller "ikke godkendt". Dog kan vi i enkelte grænsetilfælde vælge at give mulighed for genaflevering. Genafleveringsfristen er mandag d. 15. december kl. 10:00.

I første del-opgave fik I lov at øve jer lidt. Nu har I fået respons på den første del-opgave. Denne respons forventes I at tage hensyn til i besvarelsen af anden del-opgave. Selvom man evt. har fået

godkendt første del-opgave uden helt at leve op til kravene i opgave-formuleringen, betyder det altså ikke, at anden del-opgave vurderes med samme milde øjne.

7 Yderligere formalia

Den obligatoriske opgave er et individuelt eksamensprojekt. En overtrædelse af dette er eksamenssnyd og vil blive behandlet som sådan. Man har pligt til selv at beskytte sine noter og filer mod læsning af andre. Dette kan f.eks. gøres med

```
chmod 700 <opgavedirectory>
```

Begge parter involveret i en eventuel plagiering kan blive holdt ansvarlige.

Programmet skal kunne køre på IMADAs maskiner. Man må gerne lave det hjemme, men det er helt på eget ansvar. Tekniske problemer derhjemme er ingen undskyldning. Man er også helt selv ansvarlig for, at ens filer kan flyttes uden problemer.

Der SKAL afleveres til tiden. Undtagelser herfra kræver særlige omstændigheder, som vil kunne holde i studienævnet. Vi har pligt til at behandle jer ens. Derfor: check i god tid, at jeres program kan køre på IMADAs maskiner og bestå DM02check uden fejl.

Det er dit eget ansvar at afhente din rettede besvarelse og undersøge, om besvarelsen er godkendt, eller du eventuelt skal genaflevere.

God fornøjelse!

DM02 eksamen, Efteråret 2003
Obligatorisk Opgave — Anden Del

Skriv tydeligt (maskinskrift eller blokbogstaver)

Navn:

Fødselsdato:

Brugernavn (login):

Instruktor	Fiona	Frederik	Martin	Steffen
Sæt kryds				

Besvarelsen omfatter nummererede sider.

**Kvittering for aflevering af
obligatorisk opgave, anden del, i DM02, efteråret 2003**

Udfyldes inden afleveringen

Navn:

Fødselsdato:

Brugernavn (login):

Udfyldes af sekretariatet

Modtaget den kl. af

(dato) (klokken) (initialer)