

Institut for Matematik og Datalogi
Syddansk Universitet

Afleveringsopgave 4 — DM534 efterår 2014

Dette er fjerde afleveringsopgave i DM534. Deadline er

Tirsdag den 18. november, 2014, kl. 08:15.

Opgaven skal skrives i L^AT_EX, men du behøver ikke at inkludere L^AT_EX-koden denne gang. Du kan skrive på dansk eller på engelsk. Skriv dit navn, holdnummer og navnet på din instruktør (Magnus Gausdal Find eller Christian Kudahl) på første side af afleveringen.

Du skal aflevere én pdf-fil via “SDU Assignment” på Blackboard-siden for DM534. Husk at aflevere under det korrekte holdnummer og at gemme kvitteringen. Bemærk at Blackboard lukker for afleveringen ved udløb af deadline.

Opgaven udgør en del af eksamen i DM534, så samarbejde om at udarbejde besvarelsen, kopiering fra medstuderende, internettet eller andre steder, samt andre former for brug af andres indsats er derfor eksamenssnyd. Hvis du har spørgsmål til opgaven, så kontakt Joan Boyar, Rolf Fagerberg eller din instruktør i DM534.

Opgaven skal godkendes for at du kan bestå DM534. Hvis du ikke får din første aflevering af den godkendt, eller ikke får den afleveret til deadline, vil den tælle som én af de i alt to genafleveringer, du kan lave i DM534. Din genaflevering af opgaven skal så godkendes i første forsøg.

Afleveringsopgave 4

1. I denne opgave lader vi mængder være repræsenterede ved sorterede lister. F.eks. vil mængden $\{14, 27, 13, 9, 32\}$ være repræsenteret ved listen $[9, 13, 14, 27, 32]$. Da mængder per definition er uden dubletter, er listerne det også. Dette er kendt fra træningsfasen (eksaminatorietimerne).

Beskriv i pseudokode en algoritme til at finde $A \oplus B$ ud fra to mængder A og B . Her betyder $A \oplus B$ symmetrisk mængdedifferens, som er de elementer der indgår i enten A eller B , men ikke i begge. Et eksempel:

$$\{14, 27, 13, 9, 32\} \oplus \{14, 15, 9, 32\} = \{27, 13, 15\}$$

Både input A , B og output $A \oplus B$ af algoritmen skal være sorterede lister uden dubletter. Algoritmen skal køre i tid $\Theta(|A| + |B|)$, og skal være baseret på en generalisering af algoritmen til at merge to sorterede lister til én sorteret liste. Pseudokoden skal skrives som gjort til eksaminatorietimerne, dvs. under brug af operationerne `readNext`, `isEndOfFile` og `writeNext` (samt `open` og `close`) fra API'et for sekventiel tilgang til data (se evt. eksempel på side 9 i tidligere slides på <http://www.imada.sdu.dk/~joan/intro/13slide9.pdf>).

2. Vi ser her på en hashtabel med 11 pladser (buckets), dvs. den består af et array af længde 11 og en hashfunktion som afbilder nøgler til array-indeksler mellem 0 og 10. Vi antager i denne opgave at nye nøgler afbildes til en tilfældig af de 11 pladser. Angiv svaret, inkl. beregninger, på flg. spørgsmål:
 - (a) Hvad er sandsynligheden for at der er en eller flere kollisioner efter tre indsættelser?
 - (b) Hvor mange elementer skal indsættes for at det bliver mere sandsynligt at have kollisioner end ingen kollision at have?
3. Besvar opgave 15 c fra side 433. Bemærk at opgaven beder om relationelle udtryk, ikke om et SQL-udtryk.