

## **Obligatorisk oppgave**

INF5340 Algoritmer i bioinformatikk - Vår 2006

Utdelt 7. mars 2006. Innleveringsfrist: 2. mai 2006 kl 1500.

Oppgaven må bestå for at man skal få lov til å gå opp til eksamen. Oppgaven skal løses individuelt. Oppgaven teller ikke på eksamenskarakteren. Se vedlagte erklæring.

For å få godkjent oppgaven må lesbar kildekode leveres inn og systemet demonstreres med ukjente sekvenser. Det kan implementeres i et fritt valgt programmeringsspråk.

Eksempelfiler, matriser etc vil bli lagt ut her: <http://dna.uio.no/inf5340/>

### **Oppgave 1: Optimal lokal parvis sekvenssammenstilling**

I denne oppgaven er målet å implementere en algoritme for optimal lokal sammenstilling av to proteinsekvenser, også kjent som Smith-Waterman-algoritmen. Gapstraffene skal være av den affine typen, slik at gapstraff =  $g(x) = ax + b$ , der  $x$  er lengden av gappet. Gotoh laget en forbedret utgave av denne SW-algoritmen for denne typen gapstraff med en kjøretid som er  $O(MN)$  der  $M$  og  $N$  er lengden av sekvensene.

I tillegg til å beregne og vise sammenstillingsscoren skal også selve sammenstillingen vises slik det er vanlig å vise den, inkludert posisjoner, lengde, antall identiske/lignende aminosyrer etc.

Programmet skal kunne lese inn to sekvenser fra tekstfiler i FASTA-format. Scoringsmatrisen (f.eks. BLOSUM62) skal kunne leses inn fra en fil. Gapstraff-parametrene  $a$  og  $b$  skal også kunne spesifiseres, men kan f.eks. være  $a=11$  og  $b=2$ .

### **Oppgave 2: Hurtig søkeverktøy for å finne nukleotidsekvenser i store databaser**

Oppgaven går ut på å implementere SSAHA-algoritmen som forklart på forelesninger og i utdelt artikkel: Ning *et al.* (2001) SSAHA: a fast search method for large DNA databases. *Genome Res.*, 11, 1725-1729.

Det skal lages et system som kan lese inn en nukleotidsekvensdatabase på FASTA-format. Denne databasen skal i prinsippet kunne inneholde hele menneskegenomet. Systemet skal deretter kunne lese inn forskjellige søkesekvenser (som også er nukleotidsekvenser) i FASTA-format og finne ut om og eventuelt hvor disse sekvensene har treff i databasen.

Man skal kunne spesifisere en parameter  $k$  som angir lengden på ordene som systemet skal benytte. Man skal også kunne spesifisere en parameter  $n$  som er en øvre grense for frekvensen av ord. Ord som forekommer oftere enn  $n$  ganger skal ignoreres.

Alle treff skal rapporteres. Dersom treffet strekker seg over mer enn ett ord på samme diagonal uten mismatch mellom, skal treffene slås sammen og rapporteres som ett større treff. For hvert treff skal beskrivelsen av den aktuelle databasesekvensen vises, sammen med posisjon for start og slutt på treffet både i databasesekvensen og søkesekvensen. Lengden på området som matcher skal også rapporteres, samt selve sekvensen.



# Institutt for informatikk

## **Krav til innleverte oppgaver ved Institutt for informatikk**

Ved alle pålagte innleveringer av oppgaver ved Ifi – enten det dreier seg om obligatoriske oppgaver, hjemmeeksamen eller annet – forventes det at arbeidet er et resultat av studentens egen innsats. Å utgi andres arbeid for sitt eget er uetisk og kan medføre sterke reaksjoner fra Ifis side.

Derfor gjelder følgende:

1. Hvis du tar med tekst, programkode, illustrasjoner og annet som andre har laget, må du tydelig merke det og angi hvor det kommer fra.
2. Det er greit å få hint om hvorledes en oppgave kan løses, men dette skal eventuelt brukes som grunnlag for egen løsning og ikke kopieres uendret inn.
3. Kursledelsen kan innkalle studenter til samtale om deres innlevering.

## **Gruppearbeid**

I noen kurs skal det leveres gruppearbeid. Ifi krever da at alle medlemmer av gruppen kan gjøre rede for hovedtrekkene i det innleverte arbeidet. Dessuten må alle ha utført en rimelig del av det hele, og kunne identifisere og svare i detalj for sin del.

## **Samarbeid**

Reglene om kopiering betyr ikke at Ifi fraråder samarbeid – tvert imot, Ifi oppfordrer studentene til å utveksle faglige erfaringer om det meste. Men det kreves som nevnt at man kan stå inne for det som leveres.

Hvis du er i tvil om hva som er lovlig samarbeid, kan du kontakte gruppelærer eller faglærer.

[www.ifi.uio.no/studinf/skjemaer/erklaring.pdf](http://www.ifi.uio.no/studinf/skjemaer/erklaring.pdf)

27. jan. 2004