

Instituut Applied Sciences Biologie en medisch laboratoriumonderzoek Chemie Bio-informatica

Naam:		Klas:			
Nummer:					
Soort toets:	KENNISTOETS	Uitwerking	g: Gelinieerd	papier	
Datum:	vrijdag 20 novemb	er Toetsduur	r: 90 / 115 m	inuten	
Aanvang toets:	2015 11:05				
Toetscode:	Bi6a-K		HULPMIDDELEN:		
Klas(sen):	Bi2a/Bi2b		☑ Gewone rek	enmachine	
Toets:	Informatica/Java		□ PS-kaartje		
Coursetrekker:	Bgnmh		□ Tabellenboe	ek	
Aantal vragen:	6		☐ Formule-bladen		
Aantal pagina's:	3		☐ Boeken, artikelen e.d.		
Totaal pnt.: Cesuur:	100 55 pt → 5.5		□ Voeg hulr	middel toe	
Opmerkingen van de coursetrekker: Type hier eventuele opmerkingen. Let er op dat dit voorblad op één pagina past. Haal deze tekst vervolgens weg.					
Aan te vullen door het secretariaat:					
() ()	_okaal? Vie surveilleert?		Aantal toetsen	Aantal toetsen?	

Vraag 1: Introductie datastructuren (15 pt)

Datastructuren zijn belangrijk voor diverse toepassingen. In de bioinformatica is dit extreem belangrijk.

- a) Geef een toepassing van een datatructuur of algoritme waarmee een biologische, biochemische of medische vraagstelling wordt opgelost. (5 pt)
- b) Beargumenteer dat het juist in de bioinformatica belangrijk is veel kennis te hebben van datastructuren. (5 pt)
- c) Een ArrayList en een LinkedList werken precies hetzelfde voor een programmeur toch is er op de achtergrond een verschil. Beschrijf dit verschil. (5 pt)

Vraag 2: Introductie algoritmes (10 pt)

Obama kreeg van Google CEO Eric Schmidt de vraag voorgelegd hoe je een miljoen 32-bit integers zou moeten sorteren. Hij antwoordde daarop: "I think a bubble sort is the wrong way to go".

- a) Beargumenteer waardoor dit inderdaad in de meeste gevallen niet de <u>snelste</u> manier van sorteren is. (5 pt)
- b) Geef een voorbeeld (beschrijving) van een data verzameling waarbij de bubble sort helemaal niet zo'n slechte keuze is. (5 pt)

Vraag 3: Big 0 (20 pt)

Beantwoord de volgende vragen over tijdscomplexiteit.

- a. Alle bekende genen van het menselijk genoom zijn opgenomen in een **LinkedList**. Wat is de Big O wanneer we midden in genen **inserten**? (5 pt)
- b. Alle bekende genen van het menselijk genoom zijn opgenomen in een **ArrayList**. Wat is de Big O wanneer we midden in genen **inserten**? (5 pt)
- c. Wat is de Big O bij een HashMap wanneer we een element ophalen op basis van de **waarde**? (5 pt)
- d. Alle bekende genen van het menselijk genoom zijn opgenomen in een **ArrayList**. Wat is de Big O wanneer we op bekende posities een gen **zoeken**? (5 pt)

Vraag 4: Datastructuur (25 pt)

HashMaps zijn bijzonder efficiënt in het retourneren op basis van een key. Op basis van een waarde is dit minder efficiënt (zie voorgaande vraag).

- a. Wat bedoelen we als we spreken over een irreversibele hashing? Denk hierbij bijvoorbeeld aan een MD5 checksum. (5 pt)
- b. Wanneer we de waarde ophalen in een HashMap met de sleutel is de Complexiteit O(1). Beschrijf het principe waarmee dit mogelijk is. Gebruik daarvoor ten minste de woorden buckets en hashfunctie. (10 pt)
- c. Wat gebeurt er (met de Big O) wanneer er onvoldoende buckets zijn? Geef aan waar deze naar toe zal gaan.(5 pt)
- d. Wat wordt verstaan onder "collision"? (5 pt)

Vraag 5: Datastructuur (15 pt)

Begin deze eeuw was het eerste humane genoom gesequenced en geassembleerd. Dit was een stuk eerder dan gedacht. Het was mogelijk om zo snel tot een geheel genoom te komen doordat een groot aantal kleine fragmenten (reads) tot een contig konden worden geassembleerd. Dit was mogelijk door gebruikmaking van een datastructuur.

- a. Aan welke datastructuur wordt hier gerefereerd. Ben zo specifiek mogelijk. Dus noem niet alleen de datastructuur maar ook de specificering er van. (5 pt)
- b. Teken een voorbeeld van de datastructuur met de volgende hypothetische reads: ATGGG, GCCATG, GGGCC, GTTTCA (10 pt)

Vraag 6: Compressie (15 pt)

Onderzoekers zijn er in geslaagd het humane genoom extreem te comprimeren met gebruikmaking van een referentiegenoom.

- a. Is dit een voorbeeld van lossy of lossless compressie? (5 pt)
- b. In de compressie van sequenties wordt veel gebruik gemaakt van huffman compressie. Teken de Huffman tree voor de volgende sequentie: AATTATGAATA (10 pt)