

Computational Challenges in Bioinformatics (C003711)

Wegens Covid19 kan mogelijk afgeweken worden van de onderwijs- en evaluatievormen. Dergelijke afwijkingen zullen via Ufora worden gecommuniceerd.

Cursusomvang *(nominale waarden; effectieve waarden kunnen verschillen per opleiding)*

Studiepunten 6.0 **Studietijd 180 u** **Contacturen** **60.0 u**

Aanbodsessies en werkvormen in academiejaar 2020-2021

A (semester 2)	Engels	Gent	werkcollege: geleide oefeningen	15.0 u
			zelfstandig werk	5.0 u
			project	10.0 u
			hoorcollege	30.0 u

Lesgevers in academiejaar 2020-2021

Dawyndt, Peter	WE02	Verantwoordelijk lesgever
Fostier, Jan	TW05	Medelesgever

Aangeboden in onderstaande opleidingen in 2020-2021

	stptn	aanbodsessie
Brugprogramma Master of Science in Bioinformatics (afstudeerrichting Engineering)	6	A
Master of Science in Bioinformatics (afstudeerrichting Engineering)	6	A
Master of Science in de informatica	6	A
Master of Science in de ingenieurswetenschappen: computerwetenschappen	6	A
Master of Science in Computer Science Engineering	6	A
Uitwisselingsprogramma Bioinformatics (niveau master)	6	A
Uitwisselingsprogramma informatica (niveau master)	6	A

Onderwijstalen

Engels

Trefwoorden

Performantie, rekenintensieve applicaties, big data, hardware versnelling, datastructuren & algoritmes, bio-informatica

Situering

Deze cursus gaat in op enkele computationele uitdagingen in het gebied van de bio-informatica en de technieken die aangewend worden in de wetenschappelijke literatuur om ermee om te gaan. De focus ligt hier zowel op problemen die rekenintensief zijn als problemen die dataintensief zijn ('big data' problemen). Hoewel de nadruk ligt op software oplossingen komt ook het hardware aspect aan bod. Dit opleidingsonderdeel beoogt voornamelijk de studenten die zelf geïnteresseerd zijn softwareontwikkeling voor bio-informatica als doelgroep te bereiken.

Inhoud

- Sequentiealignering
 - Needleman-Wunsch en Smith-Waterman algoritmes, geavanceerde varianten
 - Aligneren van meerdere sequenties
 - Implementatie-aspecten en hardware versnelling
- Genpredictie
 - Markov models
 - Viterbi algorithm
- String matching en applicaties
 - Datastructuren en algorithmes voor string matching: suffixbomen, suffixlijsten,

- Burrows-Wheeler transformatie, FM index
- Applicaties voor read mapping
- Graaf-gebaseerde algoritmes en applicaties
- De Bruijn grafen
- Applicatie op genome assembly

Begincompetenties

Gevorderde programmeervaardigheden in Java, Python of C/C++, gevorderde kennis van datastructuren en algoritmen

Eindcompetenties

- 1 Begrijpen van veelgebruikte datastructuren en algoritmes die frequent voor bio-informatica problemen gebruikt worden.
- 2 De belangrijkste computationeel intensieve applicaties binnen de bioinformatica begrijpen, evenals de gebruikte oplossingstechnieken (of heuristieken).
- 3 Aan de hand van de complexiteit van een algoritme en de probleemgrootte een schatting kunnen maken van de rekentijd.
- 4 Aan de hand van standaard datastructuren en ontwerpstechnieken zelf geavanceerde algoritmes en software-implementaties kunnen ontwikkelen.
- 5 Op kritische wijze algoritmes met elkaar vergelijken, met het oog op computationele haalbaarheid.
- 6 Voldoende tijd spenderen aan het ontwerp en het evalueren van de complexiteit van een algoritme alvorens tot de implementatie over te gaan.
- 7 Begrijpen van de wisselwerking tussen performantie, geheugengebruik en accuraatheid van algoritmes.

Creditcontractvoorwaarde

Toelating tot dit opleidingsonderdeel via creditcontract is mogelijk mits gunstige beoordeling van de competenties

Examencontractvoorwaarde

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

Didactische werkvormen

Hoorcollege, project, zelfstandig werk, werkcollege: geleide oefeningen

Leermateriaal

- Lesslides beschikbaar op Ufora
- Wetenschappelijke publicaties worden ter beschikking gesteld via Ufora
- Broncode voorbeelden worden via Ufora ter beschikking gesteld van de student

Referenties

- "Biological Sequence Analysis", Richard Durbin, Cambridge University Press, 1998
- "Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology", Dan Gusfield, 2001
- "Bioinformatics Algorithms: Sequence Analysis, Genome Rearrangements, and Phylogenetic Reconstruction", Enno Ohlebusch, Oldenbusch-Verlag, 2013.

Vakinhoudelijke studiebegeleiding

- Extra uitleg via e-mail.
- Bijkomende informatie via Ufora.

Evaluatiemomenten

periodegebonden en niet-periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Schriftelijk examen, openboekexamen, mondeling examen

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Schriftelijk examen, openboekexamen, mondeling examen

Evaluatievormen bij niet-periodegebonden evaluatie

Vaardigheidstest, verslag

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is niet mogelijk

Toelichtingen bij de evaluatievormen

- Periodegebonden evaluatie: openboekexamen (oefeningen), mondeling examen met schriftelijke voorbereiding
- Niet-periodegebonden evaluatie: beoordeling van huistaken (broncode + verslag)

Eindscoreberekening

Voor beide examenperiodes wordt de eindscore als volgt samengesteld: 25% huistaken, 75% examen. Voor beide examenperiodes geldt dat voor het examenonderdeel een minimumscore van 50% dient gehaald te worden. Is aan deze voorwaarde niet voldaan, dan wordt de eindscore afgetopt op 9/20 voor dit opleidingsonderdeel.