

<p>Kliment Olechnovič</p>	<p><u>Baltymų erdvių struktūrų analizės metodų kūrimas</u> Darbo tikslas - susipažinti su pagrindiniais biologinių makromolekulių erdvių (3D) struktūrų analizės metodais bei prisidėti kuriant naujus metodus, t.y. naują programinę įrangą. Konkreti užduotis bus individualiai derinama su studentu. Susidomėję rašykite adresu kliment.olechnovic@bti.vu.lt</p>
<p>Darius Kazlauskas</p>	<p><u>Cas9 ir Cas12 baltymų homologų domeninės architektūros nustatymas</u> Bakterijos ir archėjos kovoja su įsibrovėliais naudodamosi adaptyviomis gynybos sistemomis, žinomomis kaip CRISPR-Cas. CRISPR-Cas atsakas susideda iš trijų etapų: adaptacijos, raiškos ir interferencijos. CRISPR-Cas sistemos yra skirstomos į dvi klases pagal baltymų, dalyvaujančių interferencijos etape, architektūrą. Minėtą procesą I klasės CRISPR-Cas sistemose atlieka daugelio baltymų kompleksai, o II klasės – dideli multidomeniniai baltymai (Cas9/Cas12/Cas13). Cas9 ir Cas12 skiriasi nuo Cas13 tuo, kad turi į RuvC panašų domeną, kuris naudojamas taikinio DNR suskaidyti. Filogenetinė Cas9 ir Cas12 baltymų RuvC domeno analizė rodo, kad jis išsivystė iš TnpB baltymo, pastarajam įgaunant/išvystant papildomus domenus. Šiame darbe bus analizuojama Cas9 ir Cas12 baltymų homologų domeninės architektūros. Darbo pradžioje bus apžvelgiami ir palyginami automatinių domeninės architektūros nustatymo (ADAN) metodai. Vėliau vyks Cas9 ir Cas12 baltymų homologų charakterizavimas pasirinktais ADAN metodais bei baltymų domeninės architektūros vizualizacija. Daugiau info: darius.kazlauskas@bti.vu.lt</p>
<p>Andrius Merkys</p>	<p><u>CIF sintaksinio analizatoriaus išplėtimas C, Python bei Java kalboms</u> Crystallographic Information File (CIF) formatas yra plačiai naudojamas kristalografijos, chemiškųjų mokslų srityse bei už jų ribų. Tačiau didelė programinės įrangos dalis palaiko nepilną CIF formato standartą, dėl ko gali kilti įvairios failų įskaitymo klaidos (Merkys et al. 2015). „codcif“ – bendro pobūdžio CIF formato analizatorius/konstruktorius (angl. reader/writer), sukurtas siekiant pilnai palaikyti CIF standartą. Šio kursinio darbo tikslas – išplėsti „codcif“ palaikymą C, Python ir Java programavimo kalboms. Prieš pasirenkant temą susisiekti su vadovu el. paštu andrius.merkys@gmc.vu.lt.</p>
<p>Andrius Merkys</p>	<p><u>dREL programavimo kalbos kompiliatorius</u> Duomenų kokybės kontrolė siekiant užtikrinti jų atitikimą tam tikriems reikalavimams taikoma jau ilgą laiką. XML ir JSON</p>

	<p>duomenų formatus lydi schemas – programiniu būdu nuskaitytų apribojimų specifikacijos. Tačiau nepilnos pagal Tiuringą kalbos visuomet leidžia aprašyti tik dalį. Todėl buvo sukurta pilna pagal Tiuringą dREL kalba duomenų validavimui pagal jų aprašymus žodynuose (Spadaccini et al., 2012). Pagrindinė jos taikymo sritis – kristalografinių duomenų aprašymas. Kadangi dREL vis vystoma ir dar nėra plačiai paplitusi, kol kas žinomas tik vienas jos kompiliatorius/interpretatorius – PyCIFRW, parašytas Python programavimo kalba ir paskutinį kartą išleistas 2018 m. pradžioje. Šio kursinio darbo tikslas – sukurti dREL kalbos kompiliatorių/interpretatorių, kurį būtų nesudėtinga keisti ir plėsti dREL kalbai vystantis. Prieš pasirenkant temą susisiekti su vadovu el. paštu andrius.merkys@gmc.vu.lt.</p>
<p>Andrius Merkys</p>	<p><u>Kristalografinės informacijos tvarkymas ir jo automatizavimas</u></p> <p>Atviroje kristalografinėje duomenų bazėje COD veikia autonominės programos-robotai, be žmogaus-operatoriaus priežiūros nukeliantys naujausias išspręstas struktūras iš mokslinės spaudos, patikrinantys jas bei patalpinantys COD duomenų bazėje. Nemažai automatiškai nukeltų struktūrų patikrinimo stadijoje yra atmetamos dėl nesunkiai pataisomų klaidų: neteisingai įvesti skaičiai, nurodyti matavimo vienetai, „žmogaus kalba“ naudojama ten, kur prašoma griežto formato. Šio kursinio darbo tikslas – išanalizuoti sukauptus patikrinimų klaidų pranešimus, nustatyti dažniausiai pasitaikančias automatiškai pataisomas klaidas, sukurti programinę įrangą klaidoms taisyti bei šią įrangą įtraukti į automatinį robotų darbo srautą. Prieš pasirenkant temą susisiekti su vadovu el. paštu andrius.merkys@gmc.vu.lt.</p>
<p>Andrius Merkys</p>	<p><u>Pasikartojančios kristalografinės informacijos aptikimas</u></p> <p>Kuriant duomenų bazines svarbu apsaugoti nuo informacijos dubliavimo. Tačiau tai ne visada yra lengva užduotis. Šio darbo tikslas – pasikartojančių įrašų aptikimo metodų kūrimas Atvirai kristalografinėi duomenų bazei COD. Prieš pasirenkant temą susisiekti su vadovu el. paštu andrius.merkys@gmc.vu.lt.</p>
<p>Andrius Merkys</p>	<p><u>Grafinė ir programatinė sąsaja reliacinėms duomenų bazėms</u></p> <p>Reliacinės duomenų bazės labai plačiai taikomos tiek mokslinei, tiek nemokslinei informacijai laikyti bei naršyti. Atviro kodo duomenų bazių valdymo sistemos, tokios kaip MySQL ar MariaDB, yra patogios naudoti komandinės eilutės ar programatinėse aplinkose, tačiau darbą su duomenų bazėmis paspartintų grafinė naudotojo sąsaja (Web UI). Vienas iš mėginimų tai įgyvendinti yra RestfulDB projektas, kurio tolimesnis vystymas ir būtų šio darbo tikslas. Darbui atlikti</p>

	<p>reikalingos geros Perl programavimo kalbos bei interneto technologijų (HTML, CSS, JavaScript) žinios. Pageidautinos SQL žinios. Prieš pasirenkant temą susisiekti su vadovu el. paštu andrius.merkys@gmc.vu.lt.</p>
<p>Karolis Koncevičius</p>	<p><u>Individo požymių nustatymas naudojantis metilinio duomenimis</u></p> <p>Šiame darbe, naudojant DNR metilinio mikrogardelių duomenis, studentui bus siūloma pabandyti nustatyti individo amžių, lytį, etninę kilmę, rūkymo statusą, ir kitus biologinius požymius. Konkreti užduotis bus derinama individualiai. Darbui numatoma naudoti R programavimo kalbą. Modelio kūrimui - įvairius regulizuotus tiesinius ("lasso", "elastic net"), atsitiktinio miško ("random forest") ir kitus regresijos ir klasifikavimo metodus.</p>
<p>Saulius Gražulis</p>	<p>Mažų molekulių kristalų kontaktų paviršiai</p> <p>Siūloma peržvelgti visas COD organinių kristalų struktūras, visų pirma tas, kuriuose yra vaistinių medžiagų molekulės ar į jas panašios molekulės. Surasti šių molekulių kontaktus su *savo pačių* kristalais, aprašyti šių kontaktų paviršius. Surasti tas molekules, kurių kompleksai su baltymais patalptinti PDB archyve. Palyginti mažos molekulės kristalo ir baltymo kontaktinius paviršius; nustatyti, ar pagal šių paviršių panašumą galima prognozuoti susimaišymą su baltymu. Prieš pasirenkant susisiekti su vadovu el. paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>Paskirstytos, patikimos ir atsparios trikdžiams bei padalinimams COD duomenų bazės kūrimas.</p> <p>Šiuo metu COD duomenų bazė, organizuota kaip centrinis (angl. "master") serveris, kurio duomenis patikimumo dėlei replikuoja visa eilė antrinių (angl. "mirror") kompiuterių. Tokia sistema, deja, neužtikrina nenutrūkstamo sistemos darbo, centriniam serveriui išėjus iš rikiuotės ar nutrūkus ryšiui tarp centrinio serverio ir Interneto. Darbo metu bus siūloma realizuoti paskirstytą, lygių serverių mainais (angl. "peer-to-peer") paremtą sistemą, atsparią sistemos padalinimui, užtikrinančią duomenų neprieštarinumą (consistency) ir minimizuojančią sistemos prastovas. Nors CAP teorema teigia, kad visų trijų tikslų (neprieštarinumo, prieinamumo ir atsparumo padalinimams) vienu metu pasiekti neįmanoma, bus bandoma surasti inžinerinius sprendimus, leidžiančius minimizuoti prastovas (prastovas, duomenų praradimą ir pan.), atsiradusius dėl</p>

	<p>to, kad COD bus realizuota kaip paskirstyta duomenų bazė. Darbo metu reikės išnagrinėti įvairius galimus sistemos variantus, įvertinant įvairius kompromisus (prieinamumas/neprieštaringumas, prieinamumas/patikimumas ir pan.). Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>Duomenų kokybės užtikrinimas ir duomenų validavimas kristalografinėje duomenų bazėje COD.</p> <p>Duomenys mokslinėje duomenų bazėje naudingi tik tada, kai jie yra patikimi ir teisingi. Deja, net aukšto lygio mokslinėse publikacijose ne visada užtikrinamas duomenų teisingumas ir atitikimas formaliems reikalavimams. COD (Crystallography Open Database) duomenų bazės kūrėjai šiuo metu pasiekė, kad visi duomenų failai yra sintaksiškai teisingi (atitinka IUCr CIF formato reikalavimus) ir gali būti apdorojami automatiškai. Sekantis žingsnis link aukštos kokybės duomenų bazės -- semantinis duomenų patikrinimas (validacija) pagal IUCr sukurtas ontologijas -- CIF žodynus (angl. "CIF dictionaries"), ir prasminių klaidų paieška, naudojant statistinius metodus. Darbo metu bus siūloma tobulinti COD duomenų validatorių, atlikti visų duomenų validaciją, pagal validacijos pranešimus sukurti automatines klaidų taisymo priemones, ištaisyti tas semantines klaidas, kurias įmanomai vienareikšmiškai atpažinti, pažymėti nepataisomas klaidas, bei integruoti klaidų taisymo priemones į COD duomenų įkėlimo svetainę. Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>Didelės apimties duomenų archyvavimas paskirstytoje, lygių partnerių bendradarbiavimu paremtoje (angl. "peer-to-peer") duomenų saugykloje.</p> <p>Naujausios IUCr (Tarptautinės kristalografų sąjungos, angl. International Union of Crystallography) rekomendacijos siūlo archyvuoti visus pradinius duomenis, panaudotus struktūros nustatymui, įskaitant difrakcijos (išsklaidytų Rentgeno spindulių) vaizdus, užregistruotus monokristalinių difraktometrų. Šios rekomendacijos įgyvendinimas kelia naujus iššūkius -- bus reikalingas gerokai didesnis pastovios atminties (diskų, juostų) kiekis, negu naudotas iki šiol, ir duomenys turi būti prieinami bent jau ateinančius dešimtmečius, t.y. pergyventi kelias kompiuterinės įrangos kartas. Visa tai susiję su</p>

	<p>papildomomis sąnaudomis ir duomenų laikymo kaštais. Vienas iš galimų šių problemų sprendimo būdų -- panaudoti paskirstytą, daugelio institucijų ir/arba individų palaikomą duomenų archyvavimo sistemą, turinčią pakankamą duomenų perteklumą, užtikrinantį patikimą sistemos darbą ilgą laiką. Darbo metu bus siūloma išnagrinėti įvairių partnerių bendradarbiavimu“ (angl. "peer-to-peer") bei paskirstytų duomenų bazių sistemų (Gnutella, GUNet, OFFSystem, Riak, Apache Cassandra, ir t.t.) tinkamumą nurodytam tikslui ir galimai sukurti veikiantį sistemos prototipą. Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>Senų kristalografinių duomenų skaitmeninimas COD duomenų bazei.</p> <p>Dalis duomenų apie svarbius cheminius junginius, tame tarpe apie jų erdvinės struktūras, buvo publikuota prieš plačiai plintant kompiuteriams ir internetui, tad šie duomenys yra prieinami tik "popieriniame" pavidale, išbarstyti po daugelį skirtingų žurnalų ar leidinių, arba patalpinti nuosavybinėse duomenų bazėse. Tokia situacija techniškai ir/arba juridškai apsunkina duomenų radimą bei panaudojimą. Darbo metu siūloma sukurti įrankius struktūrinės informacijos įvedimui ar optiniam simbolių atpažinimui, PDF failų tekstų analizei, siekiant atpažinti ir išskirti kristalografinius duomenis, ir galimai suskaitmeninti senas publikacijas, įkeliant jų duomenis į atvirą kristalografinę duomenų bazę COD (Crystallography Open Database). Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>Vidutinių trimačių simetrijos grupių apskaičiavimas iš keturmačių moduluotų struktūrų simetrijos operatorių.</p> <p>Pastaruoju metu daugėja informacijos apie medžiagos būvį, kuris, nors ir turi daugumą kristalo savybių (pvz., sklaido Rentgeno spindulius siaurais koncertuotais atspindžiais), nėra tikras kristalas, nes negali turėti periodinės gardelės, suderinamos su stebima objekto ar sklaidymo vaizdo simetrija, tokia kaip penkto laipsnio simetrijos ašis. Tai -- kvazikristalai (http://en.wikipedia.org/wiki/Quasicrystals) ir (nebendramatės) moduluotos struktūros. Šioms struktūroms aprašyti kuriamas matematinis aparatas,</p>

	<p>panaudojantis simetrijos grupių teoriją. Pasirodo, kad neperiodines trimates struktūras galima aprašyti kaip periodinių struktūrų daugiamatėse erdvėse pjūvius. Pavyzdžiui, kai kurias moduluotas struktūras galima nagrinėti kaip periodinių 4-mačių gardelių pjūvius. Perėjimas į aukštesnių matavimų erdves leidžia panaudoti jau žinomą erdvinių simetrijos grupių mat. aparatą, ir kompaktiškai aprašyti neperiodines struktūras. Darbo metu bus siūloma sukurti programinę įrangą, kuri tikrintų keturmačių simetrijos grupių aprašymus, pagal šiuos aprašymus sukurtų vidutinius nemoduluotos trimatės simetrijos grupės aprašus, ir integruoti šiuos algoritmus į duomenų bazę COD, kad būtų galima efektyviai kaupti ir tvarkyti neperiodinių medžiagos pavyzdžių aprašymus. Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>BOINC serverio ir klientų parengimas statistiniams skaičiavimams ir jų pritraukimas COD duomenų bazės analizei.</p> <p>Statistiniai skaičiavimai, paremti Bajeso statistikos principais, duoda universalią ir koherentišką skaičiavimo metodiką, bet reikalauja itin daug skaičiavimo resursų. Vienas iš būdų tokius resursus surinkti -- panaudoti masinį paralelizmą „savanorių skaičiuotojų“ (angl. "volunteer computing") pateiktuose kompiuteriuose. Šiuo principu yra paremta Berklio universiteto BOINC sistema. Darbo metu bus siūloma: a) paleisti BOINC sistemos serverį; b) parašyti paprasčiausius BOINC klientus; c) parašyti klientus, skirtus COD atstumų ir jungčių parametrų tikimybių pasiskirstymų pasiskirstymų radimui ir atnaujinimui, naudojant Bajeso statistikos metodus, ir skaičiavimų organizavimas. Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>Gramatikos atstatymas iš kalbos pavyzdžių.</p> <p>Gera žinoma yra uždaviniai pagal nurodytą gramatiką nustatyti, ar duota simbolių eilutė priklauso gramatikos generuojamai kalbai, ir ar dvi gramatikos yra ekvivalenčios; žinomi efektyvūs šių uždavinių sprendimo būdai daugeliui praktiškai svarbių kalbų klasių. Tačiau praktikoje kartais tenka spręsti atvirkščią uždavinį: pagal kalbai priklausančių ir nepriklausančių eilučių pavyzdžius sukonstruoti minimalią gramatiką, generuojančią tokią</p>

	<p>kalbą. Šis uždavinys kur kas blogiau apibrėžtas (neturi unikalios sprendimo) ir efektyvūs sprendimo būdai bendru atveju nėra žinomi. Darbo metu bus pasiūlyta suformuluoti ir išspręsti uždavinį paprasčiausiai -- reguliarių kalbų -- klasei. Konkrečiai, pagal duotas eilutes su teisingais ir klaidingais duomenų (teksto) pavyzdžiais reikės sukonstruoti reguliarias išraiškas, kurias atitiktų teisingos eilutės bet neatitiktų klaidingos eilutės. Galimi sprendimo būdai būtų euristikos, kodo evoliucija ir genetiniai algoritmai, apmokomų neuronų tinklų ar atraminių vektorių mašinų panaudojimas. Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>Teorinės kristalografinės duomenų bazės TCOD duomenų validavimas.</p> <p>Pastaruoju metu labai sparčiai vystosi skaičiuojamosios chemijos metodai, leidžiantys suskaičiuoti kristalų bei molekulių struktūras naudojant pamatinius kvantinės mechanikos principus. Atsiranda vis daugiau laisvų programų, leidžiančių atlikti šiuos skaičiavimus, ir sparčiai daugėja suskaičiuotų struktūrų duomenų. Atsiranda poreikis šiuos skaičiavimo rezultatus sistematizuoti, įvertinti jų patikimumą bei palyginti su eksperimentiniais duomenimis. Tuo tikslu buvo paleista TCOD duomenų bazė. Darbo metu bus pasiūlyta įgyvendinti duomenų kokybės patikrinimo programas ir palyginti skirtingas suskaičiuotas struktūras tarpusavyje ir su eksperimentiškai nustatytomis struktūromis. Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Irus Grinis</p>	<p>Baltymų struktūrų nustatymas naudojant gilųjį mokymąsi.</p> <p>Ši tema gana sudėtinga, bet iš kitos pusės labai perspektyvi. Pradėti galima nuo kokio nors kurselio apie gilųjį mokymąsi (pvz. https://www.tensorflow.org/tutorials). Kitas žingsnis - susipažinti su įranga https://github.com/deepmind/deepmind-research/tree/master/alphafold_casp13 ir https://yanglab.nankai.edu.cn/trRosetta/. Trečias ir kiti žingsniai - „eksperimentavimas“ su mineta įranga.</p>
<p>Irus Grinis</p>	<p>Evoliuciniai skaičiavimai ir algoritmai, jų taikymas gamtos ir tiksliuosiuose moksluose</p> <p>Ši tema labai plati. Pradžioje temą pasirinkusiam studentui reikės „apsižvalgyti“. Galima pradėti nuo liaudiško šaltinio. Rimtesnis ir platesnis šaltinis yra monografija. Po to galima</p>

	<p>būtų susikoncentruoti ties 3-4 algoritmų, iširti jų veikimo rezultatus konkrečioms duomenims, palyginti tarpusavyje jų efektyvumą, pasiūlyti konkrečias jų taikymų rekomendacijas.</p>
Irus Grinis	<p>Bioinformatikos svetainės priežiūra ir plėtimas.</p> <p>Mūsų portalas gyvuoja jau metus. Jo pagrindinė paskirtis bioinformatikos mokslo ir studijų populiarinimas. Numatoma, kad kiekviena bioinformatikų laida turės savo atstovą - redaktorių - užsiimantį ne tik svetainės priežiūra, bet ir ruošiantį naujus straipsnius, mokomąją medžiagą ir kitus resursus minėtam portalui.</p>
Irus Grinis	<p>Rekurentiniai neuroniniai tinklai ir jų taikymas biomoksluose.</p> <p>Rekurentiniai neuroniniai tinklai plačiai taikomi įvairiose srityse. Šio darbo tikslas -- susipažinti su kai kuriais esamais RNN taikymais biosekų analizėje ir iširti galimybes plačiau panaudoti juose natūraliosios kalbos apdorojimo technikas. Pradžiai reikėtų susipažinti aplamai su įrankiais pvz. https://www.tensorflow.org/, apžvelgti turimas publikacijas (pvz. https://doi.org/10.1093%2Fbioinformatics%2Fbtm247, https://doi.org/10.1038/s41598-019-52196-4)</p>
Irus Grinis	<p>Edukaciniai žaidimai gamtos ir tiksluosiuose moksluose.</p> <p>Nors kompiuterinių žaidimų industrija skaičiuoja kelis dešimtmečius, bet kai kalba užėina apie platesnį jų taikymą mokyme, atsiranda nemažai problemų. Šio darbo tikslas - pabandyti sukurti kokį nors paprastą edukacinių žaidimų kūrimo įrankio prototipą, kuriuo galėtų naudotis gamtos ir tikslųjų mokslų mokytojai/dėstytojai</p>
Irus Grinis	<p>Web aplikacijų kūrimas naudojant Biopython.</p> <p>Su Biopython'u esame susipažinę per Bioinformatikos II kursą. Šios darbo tikslas būtų parinkti python web karkasą (pvz. Django, Flask, ...) ir integruoti jį su Biopython ir kai kuriais kitais python kalba paremtais biomedicininiais įrankiais.</p>
Gintautas Bareikis	<p>Genetiniai algoritmai ir jų taikymas modeliuojant biosistemas.</p> <p>Mus supanti aplinka yra unikali, o šią aplinką įtakojantys veiksniai jei ne vienodi, tai bent jau panašūs. Norėdami nustatyti, kokį poveikį biosistemai daro įvairūs veiksniai, mes tai galime atlikti kurdami realaus pasaulio modelius bei jų aplinką. Gamtinių procesų negalime kartoti "atsukdami laiką", tuo tarpu</p>

	<p>kompiuterinis modeliavimas sudaro prielaidas žymiai platesnei biosistemų analizei, stebint ją esant įvairiems poveikiams, gražinant sistemą prie bet kurios išeities būsenos. Darbo tikslas - kurti biosistemas ir modeliuoti jų vystymąsi priklausomai nuo aplinkos sąlygų.</p>
<p>Kotryna Kvederavičiūtė</p>	<p>Mažųjų RNR analizė ir tyrimai Mažosios nekoduojančios RNR yra vieni iš genų raiškos reguliatorių (potransliaciniame lygyje). Šios RNR yra labai heterogeniškos tiek savo dydžiu, tiek struktūra ir tik nedidelės jų dalies funkcija/funkciniai mechanizmai yra žinoma. Darbo metu bus siūloma atlikti mažųjų RNR analizę. Konkretūs atliekami darbai priklausys nuo to, kokie duomenys tuo metu bus prieinami laboratorijoje ir gali apimti mažųjų RNR sekoskaitos duomenų analizę, mažųjų RNR savybių analizę ir įvairios programinės įrangos, skirtos dirbti su mažosiomis RNR kūrimą. Prieš pasirinkdami temą, studentai turėtų susisiekti su temos vadovu (kotryna.kvederaviciute@gmail.com).</p>