

<p>Kliment Olechnovič</p>	<p>Baltymų erdviųjų struktūrų analizės metodų kūrimas Darbo tikslas - susipažinti su pagrindiniais biologinių makromolekulių erdviųjų (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>Cas9 ir Cas12 baltymų homologų domeninės architektūros nustatymas 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>Karolis Koncevičius</p>	<p>Individo požymių nustatymas naudojantis metilinimo duomenimis Šiame darbe, naudojant DNR metilinimo 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 Siūloma peržvelgti visas COD organinių kristalų struktūras, visų pirma tas, kuriuose yra vaistinių medžiagų molekulės ar į jas</p>

	<p>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>Polimerų aptikimas kristalografinėje duomenų bazėje COD</p> <p>Krisaluose molekulės gali sudaryti begalines persipynusias grandines – polimerus. Tokios medžiagos labai intensyviai tiriamos pastaraisiais metais; tai metaloorganiniai karkasai, organiniai karkasai, ceolitai. Polimerinės molekulės turi daug įdomių savybių, priklausanšių nuo jų polimerinės sandaros: jos „kvepuoja“ (keičia savo tūrį, neprarasdamos struktūros), į tuščias ermes galima įterpti kitas molekules. Ieškant tokių molekulių su įdomiomis savybėmis, reikia patikti visas tokias struktūras ir jas suklasifikuoti. Tai ir bus siūloma šiame darbe; reikės sukurti polimerų aprašymo metodus naudojant faktorgrafų teoriją, patobolinti esamą programinę įrangą ir atlikti skaičiavimus su visa COD duomenų baze, rezultatus pateikiant Web svetainėje ir mašina skaitomu pavidalu per REST sąsają. Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>Topologinių metodų taikymas kristalografijoje: mazgų paieška kristalų struktūrose</p> <p>Molekulės medžiagose gali sudaryti įvairias netikėtas struktūras: kelios molekulės gali būti sukabintos, kaip žiedai grandinėje (katenanai); molekulės gali būti steriškai sukabintos viena su kita, bet tuo pat metu laisvai judėti viena kitos atžvilgiu (rotaksanai). Teoriškai įmanomi tokių sukabintų žiedinių molekulių polimerai. Šio darbu uždavinys būtų aptikti visas panašaus pobūdžio struktūras atviroje kristalografinėje duomenų bazėje COD. Prieš pasirenkant susisiekti su vadovu el paštu, adresu: grazulis@ibt.lt.</p>
<p>Saulius Gražulis</p>	<p>Kristalografinių ir bioinformatinių algoritmų formalus verifikavimas</p> <p>Programinės įrangos klaidos turi daugybę neigiamų paskemių – nuo erzinančio laiko šavistymo iki atšauktų mokslinių straipsnių iš gerai žinomų mokslo žurnalų; o blogiausiu atveju klaidingos programos kritiškose sistemose gali kainuoti žmonių gyvybes. Testavimas, nors ir padeda pakelti programų kokybę, negali įrodyti, kad programos yra teisingos; jis tik gali parodyti esamas</p>

	<p>programų klaidas. Pastaruoju metu intensyviau vystomi formalaus programų verifikavimo metodai, leidžianti įrodyti kaip matematinę teoremą, kad programa atitinka specifikaciją. Pagrinde šie metodai taikomi itin svarbioms sistemoms, kurių neteisingas veikimas turėtų labai blogų pasekmių – geležinkelių valdyme, avionikoje, pramoninių procesų valdyme. Bet, plintant formalaus įrodymo įrankiams, pravartu būtų ir pagrindinams bioinformatikos bei kristalografijos algoritmams pritaikyti formalaus įrodymo priemones. Darbe bus siūloma įgyvendinti keletą bioinformatinių algoritmų (sekų palyginimo, FFT) Ada programavimo kalba ir įrodyti jų teisingumą SPARK sistemos pagalba. 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</p>

	<p>kompiuterinės įrangos kartas. Visa tai susiję su 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 veikiančią 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 apsunkena 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, 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</p>

	<p>programinę įrangą, kuri tikrintų keturmačių simetrijos grupių aprašymus, pagal šiuos aprašymus sukurtų vidutinius nemoduliuotos 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 pritraukymas 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 parametru 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>Gerai žinomi 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 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>Tikimybinis programavimas ir jo taikymai</p> <p>Tikimybinis programavimas yra viena iš sistemų mokymosi ir dirbtinio intelekto sričių. Joje bandoma sukurti stebimos sistemos modelį, kuriame mes tam tikrų parametrų nežinome, bet bandome juos „atspėti“ iš sistemos „raiškos“. Tikimybiniam programavimui sukurta nemažai specializuotų kalbų/aplinkų, pvz. Gen.jl (Julia kalba parašyta aplinka), Anglican (integruotas su funkcinė kalba Closure), PyMC3 (Python pagrindu sukurtas modulis) ir t.t. Kursinio darbo metu pradžioje reikėtų pasirinkti kokią nors iš išvardintų kalbų, susipažinti su ja, pagaliau pabandyti pritaikyti įgytus įgūdžius konkrečioms duomenims, pvz. Iš https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7904853/ ar pan.</p>
<p>Irus Grinis</p>	<p>Evoliuciniai skaičiavimai ir algoritmai, jų taikymas gamtos ir tiksluosiuose moksluose</p> <p>Ši tema labai plati. Pradžioje temą pasirinkusiam studentui reikės „apsižvalgyti“. Galima pradėti nuo liaudiško šaltinio. Rimtesnis ir platesnis šaltinis yra monografija. Tai pat galima pažiūrėti bioinformatikai „artimą“ apžvalgą https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3963368/ Po to galima bus pasirinkti tam tikrą konkretų optimizavimo uždavinį ir pritaikyti jam įgytus įgūdžius.</p>
<p>Irus Grinis</p>	<p>Bioinformatikos svetainės priežiūra ir plėtimas</p> <p>Mūsų portalas gyvuoja jau trečius metus. Jo pagrindinė paskirtis bioinformatikos mokslo ir studijų populiarinimas. Numatoma, kad kiekviena bioinformatikų laida turės savo atstovą - redaktorių - užsiimančią ne tik svetainės priežiūra, bet ir ruošiančią naujus straipsnius, mokomąją medžiagą ir kitus resursus lietuvių ir anglų kalbomis minėtam portalui.</p>

<p>Irus Grinis</p>	<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. Pradžiai reikėtų susipažinti apšiamai su įrankiais pvz. https://www.tensorflow.org/, apžvelgti kai kurias publikacijas (pvz. https://pubmed.ncbi.nlm.nih.gov/33315571/)</p>
<p>Irus Grinis</p>	<p>Edukaciniai web žaidimai gamtos ir tiksluosiuose moksluose</p> <p>Nors kompiuterinių žaidimų industrija skaičiuoja kelis dešimtmečius, bet kai kalba užeina apie platesnį jų taikymą mokyme, atsiranda nemažai problemų. Šio darbo tikslas: naudojant kokį nors web karkasą (pvz. Django) ir žaidimo varikliuką (pvz. Three.js, Phaser ar pan., žr. pvz. https://blog.logrocket.com/top-6-javascript-and-html5-game-engines/), pabandyti sukurti kokį nors paprastą edukacinių žaidimų kūrimo įrankio prototipą, integruoti jį pvz. Moodle VMA ar WordPress CMS taip, kad juo potencialiai galėtų naudotis gamtos ir tikslųjų mokslų mokytojai/dėstytojai.</p>
<p>Irus Grinis</p>	<p>Biologinių web aplikacijų kūrimas naudojant Django ir/ar Flask karkasus</p> <p>Django ir Flask yra populiarūs web aplikacijų kūrimo Python karkasai. Darbo metu reikėtų susipažinti su web aplikacijų su atitinkamu karkasu kūrimo pagrindais, pvz. https://docs.djangoproject.com/en/4.0/intro/tutorial01/ . Po to galima bus pabandyti atrinkti bioinformatikinius įrankius, kuriems galima būtų suprojektuoti ir sukurti atitinkamą apvalkalą.</p>
<p>Irus Grinis</p>	<p>Gilieji neuroniniai tinklai ir jų taikymas</p> <p>Tai labai plati tema, nes giliųjų DNN taikymai mūsų laikais pasitaiko „visur“. Studentui, pasirinkusiam šią temą, pirmiausia teks susipažinti su neuroniniais tinklais ir giliaja architektūra. Tam yra sukurta labai daug mokomosios medžiagos žiniatikyje (pvz.: https://www.tensorflow.org/tutorials/images/cnn). Antras žingsnis būtų konkrečios „specializacijos“, t.y. giliųjų DNN taikymo srities pasirinkimas ir tyrimų vykdymas. Turima omenyje, kad visame šiame procese dalyvaus ir vadovas :).</p>
<p>Gintautas Bareikis</p>	<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 kompiuterinis modeliavimas sudaro prielaidas žymiai platesnei biosistemų analizei, stebint ją esant įvairiems poveikiams, grąžinant sistemą prie bet kurios išėties būsenos. Darbo tikslas - kurti biosistemas ir modeliuoti jų vystymąsi priklausomai nuo aplinkos sąlygų.</p>
<p>Kotryna Kvederavičiūtė</p>	<p>tbx5 ir tcf21 genų įtakos regeneracijai tyrimai (temą gali rinktis 2 studentai). TCF21 ir TBX5 yra transkripcijos faktoriai, vaidinantys svarbų vaidmenį širdies ir kai kurių kitų organų/sistemų vystymėsi. Kursinio darbo metu būtų siekiama naudojantis viešai prieinamas pelės ChIP-seq analizių rezultatais identifikuoti pelės tbx5/tcf21 taikinių homologus zebbrinėje danijoje ir iš identifikuotų homologų atrinkti tuos, kurie yra svarbūs regeneracijai (remiantis viešais RNA-seq/microarray duomenimis). Kursinis darbas būtų atliekamas VU GMC EGIS Dr. D. Balčiūno grupėje. Prieš pasirinkdami temą studentai turėtų susisiekti su vadovu (kotryna.kvederaviciute@gmail.com).</p>
<p>Kotryna Kvederavičiūtė</p>	<p>TOP-seq sekoskaitos taikymai ir analizė TOP-seq yra vienos bazės poros rezoliucijos metodas, skirtas identifikuoti nemetilintus citozinius, 5-hidroksi/5-formil/5-karboksi citozinius. Šios citozino modifikacijos vaidina ypatingai svarbų vaidmenį epigenetiniame perprogramavime, vystymesi ir kt., bet šiuo metu daugiausiai yra ištirta ir žinoma tik apie 5-metilcitozinų įtaką ir funkcijas, o kitų modifikacijų nepriklausomos funkcijos vis dar nėra plačiai ištirtos. Kursinio darbo tikslas būtų susipažinti su TOP-seq sekoskaita ir jos praktiniu taikymu tiriant įvairias citozino modifikacijas. Tolimesni (po kursinio) darbai apimtų bendrosios TOP-seq sekoskaitos arba vienos ląstelės TOP-seq sekoskaitos duomenų analizę ir metodo vystymą. Kursinis darbas būtų atliekamas VU GMC DMTS Dr. E. Kriukienės grupėje. Prieš pasirinkdami temą studentai turėtų susisiekti su vadovu (kotryna.kvederaviciute@gmail.com).</p>