Teoria: Bioinformatica este studiul structurii informației biologice și a sistemelor biologice. Ea aduce împreună datele de cercetare gnomică cu teoria și instrumentele matematicii și a științei calculatoarelor.

Bioinformatica implică aplicarea principiilor teoriei, experimentelor și designului în contextul analizei datelor biologice.

Experimentul: În bioinformatică, experimentul se referă la colectarea și generarea datelor biologice, cum ar fi secvențele de ADN, secvențele de proteine, expresia genelor sau datele metabolomice. Aceste date sunt obținute prin metode experimentale precum secvențierea ADN-ului, analiza de microaray, spectrometrie de masă și alte tehnici biologice. Experimentele generează seturi mari de date care sunt ulterior analizate și interpretate folosind tehnici și algoritmi bioinformatici.

Design-ul: Design-ul bioinformaticii se referă la procesul de creare și implementare a metodelor și algoritmilor pentru a rezolva problemele specifice din domeniul bioinformaticii. Acesta implică dezvoltarea de noi modele matematice și statistice, conceperea de algoritmi eficienți și crearea de instrumente software pentru analiza datelor biologice. Design-ul bioinformaticii se concentrează pe optimizarea performanței, acurateței și interpretării rezultatelor analizei datelor biologice.

Probleme:

Alinierea secvențelor: Una dintre problemele fundamentale în bioinformatică este alinierea secvențelor biologice, cum ar fi secvențele de ADN sau proteine. Scopul este de a identifica regiunile similare sau omoloage între secvențe și de a obține înțelegerea asupra evoluției și funcției acestora.

Analiza datelor single-cell: Tehnologiile de secvențiere single-cell permit analiza celulelor individuale, dezvăluind heterogenitatea și dinamica acestora în cadrul țesuturilor și organismelor. Problemele deschise includ identificarea subtipurilor celulare, reconstrucția traiectoriilor dezvoltării celulare și integrarea datelor single-cell cu datele genetice și funcționale.

Identificarea sabloanelor (motif discovery)

Contextul problemei:

Genele sunt activate (exprimate) sau inactivate de către anumite proteine reglatoare; genele sunt considerate activate (exprimate) când produc un efect (de exemplu încep să sintetizeze proteine care distrug agentul patogen prezent în cazul unei infecții). Proteinele reglatoare se “leagă” în regiuni specifice ale genelor determinând stimularea sau inhibarea ARN polimerazei (controlând astfel procesul de transcripție). Aceste regiuni specifice sunt de regulă secvențe ADN scurte, numite motiv sau șablon (de exemplu la musculița de oțet genele implicate în reacțiile imunologice au la început o secvența de forma: TCGGGGATTTCC). Identificarea aceluiași șablon în mai multe gene cu caracter reglator sugerează prezența unei relații între acestea.

Enuntul problemei:

Să se identifice șabloane frecvente (motive) într-un set de secvențe AND

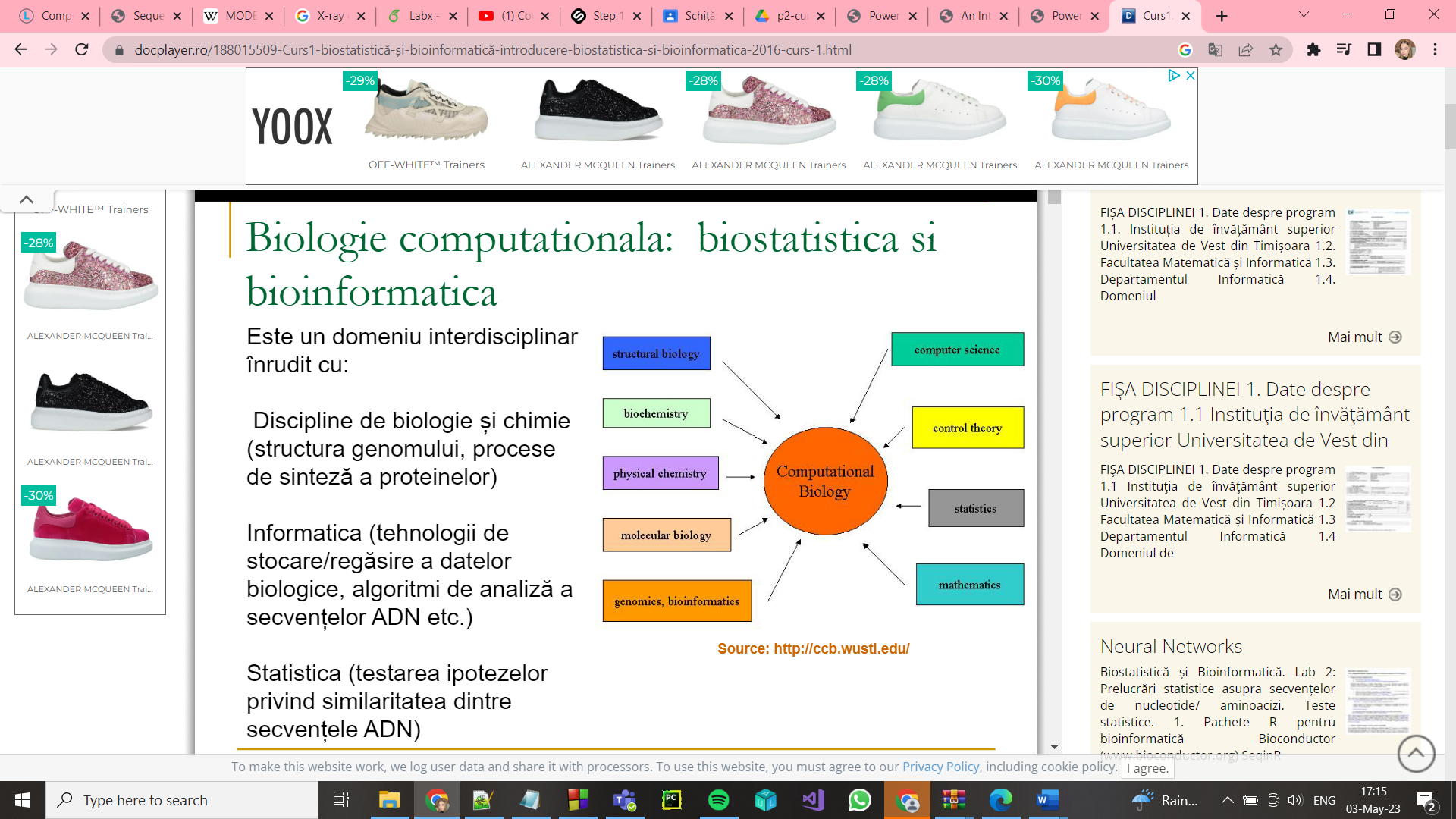
Abordare:

* Se determină frecvențe ale succesiunilor de litere (ex: 2, 3 litere)
* Identificarea secvențelor frecvente (de exemplu “the” în engleză) permite efectuarea unor inferențe asupra simbolurilor necunoscute
* In baza acestei analogii rezultă că, la fel cum cunoașterea frecvenței cuvintelor/simbolurilor dintr-o limbă permite decodificarea unui text codificat, cunoașterea “distribuției” asociate unei secvențe poate fi utilă in “descifrarea ADN-ului”
* Aceasta motivează analiza frecvențelor unor șabloane în secvențele de nucleotide

Relații cu alte discipline

Bioinformatica este un capitol de sine stătător al informaticii medicale. Însă, prin specificul obiectului său de a urmări modul în care este reprezentată informația în structurile vii, obiectul său este strâns legat de o serie de discipline privind structura și funcția structurilor din materia vie. Astfel, sunt necesare cunoștintele de biofizică, biochimie, biologie celulară și moleculară precum si genetică.

Din punct de vedere al abaterilor formale, legătura cu teoria probabilităților, biostatistica, teoria sistemelor și chimia cuantică este evidentă. De asemenea, se folosesc diverse limbaje de programare - R, PERL, C++, Java.



Bioinformatica a început să se dezvolte la începutul anilor 1970. A fost inițial considerată știința analizei proceselor informatice ale diferitelor sisteme biologice. În această perioadă, cercetarea în domeniul **inteligenței artificiale** folosea modele de rețea ale creierului uman pentru a genera noi algoritmi. Această utilizare a datelor biologice pentru a dezvolta alte domenii a împins cercetătorii biologici să revină la ideea utilizării computerelor pentru a evalua și compara seturi mari de date. Cantitatea de date partajate între cercetători a început să crească exponențial până la sfârșitul anilor 1980. Aceasta a necesitat elaborarea de noi metode computaționale pentru a analiza și interpreta rapid informațiile relevante.

Anatomia computațională se axează pe cercetarea cantitativă și modelarea variabilității formelor anatomice. Aceasta implică dezvoltarea și aplicarea metodelor matematice, statistice și de analiză a datelor pentru modelarea și simularea structurilor biologice și combină elemente de **anatomie, matematică aplicată dar și teoretică, învățare automata, științe computaționale, imagistică biologică, probabilitate și statistică**. Interpretarea sa folosește metadate derivate din modalitățile originale de obținere ale imaginilor senzoriale (dintre care Imagistica de rezonanță magnetică este un exemplu). Se concentrează asupra structurilor anatomice înregistrate, mai degrabă decât asupra dispozitivelor medicale de imagistică. Se aseamănă în acest sens cu **lingvistica computațională**, o disciplină care se concentrează mai degrabă pe structurile lingvistice decât pe senzorul care acționează ca mijloc de transmisie și comunicare.

**Biomodelarea computațională** este un domeniu referitor la construirea modelelor computerizate ale sistemelor biologice. Biomodelarea computațională urmărește dezvoltarea și utilizarea simulărilor vizuale pentru a evalua complexitatea sistemelor biologice. Acest lucru se realizează prin utilizarea algoritmilor specializați și a software-ulilor de vizualizare. Aceste modele permit predicția modului în care sistemele vor reacționa în medii și condiții diferite. Acest lucru este util pentru a determina dacă un sistem este robust. Un sistem biologic robust este acela care “își menține starea și funcțiile împotriva perturbațiilor externe și interne”  (Kitano, 2002), care este o caracteristică esențială pentru supraviețuirea unui sistem biologic. Biomodelarea computațională generează o arhivă amplă a acestor date, permițând accesul și analiza de către mai mulți utilizatori. În timp ce tehnicile actuale se concentrează asupra sistemelor biologice mici, cercetătorii lucrează la abordări care vor permite analiza și modelarea rețelelor mai mari. Acest lucru va fi esențial în dezvoltarea abordărilor medicale moderne în vederea creării de noi medicamente și terapii genetice.

O altă arie de studiu în care poate fi aplicată metoda computațională este este **biologia cancerului**, cu scopul de a determina viitoarelor mutații ale cancerului printr-o abordare algoritmică a analizei datelor (Yakhini&Jurisica, 2011). Cercetarea în acest domeniu a condus la utilizarea măsurătorilor de mare viteză a unui volum mare de informații. Măsurarea unui debit mare de date presupune colectarea a milioane de puncte de date și este posibilă utilizând robotică și alte dispozitive de detectare. Aceste date sunt colectate din ADN, ARN și alte structuri biologice. Domeniile de interes includ determinarea caracteristicilor tumorilor, analizarea moleculelor care pot fi deterministe în producerea cancerului și înțelegerea modului în care genomul uman se raportează la cauzalitatea tumorilor și a cancerului.

**Neuropsihiatria computațională** este un nou domeniu în curs de dezvoltare care utilizează modelarea matematică și computațională a mecanismelor creierului implicate în tulburările psihice. S-a demonstrat deja prin mai multe inițiative că modelarea computațională este o contribuție importantă pentru înțelegerea circuitelor neuronale care ar putea genera funcții mentale și disfuncții precum schizofrenia de exemplu.

Neuroștiința computațională presupune studiul funcționării creierului din punct de vedere al proprietăților de procesare a informației din diferitele structure care compun sistemul nervos. Este o ramură a neuroștiinței pure, și își propune să analizeze date legate de creier cu scopul de a crea aplicații practice. Acest lucru implică utilizarea de modele ale creierului pentru a examina anumite aspecte specifice ale sistemului neurologic. Din multele feluri de modele, cele mai importante sunt cele realistice si cele simplificate. Modelele realistice incearcă sa reprezinte fiecare aspect al creierului, incluzând cât mai multe detalii cu putință, și un număr foarte mare de variabile. Aceste modele oferă o cantitate foarte mare de informații despre creier, dar au și cea mai mare predispoziție la erori, întrucât nu țin cont de părțile de structură celulară depre care cercetătorii încă nu știu. Modelele realistice sunt cele mai grele de realizat din punct de vedere computational și foarte dificile de implementat. Pe de cealaltă parte, modelele simplificate își propun să reducă dimensiunea modelului pentru a examina o proprietate fizică cât mai specific a sistemului neurologic. Acest lucru permite rezolvarea problemelor computaționale mai intense și reduce potențialul de eroare a unui model realistic, întruct sunt mult mai puține necunoscute luate în calcul. Cercetarea în domeniul neuroștiinței computaționale duce la îmbunătățirea algoritmilor folosiți și a structurilor de date pentru a crește viteza acestor calcule.

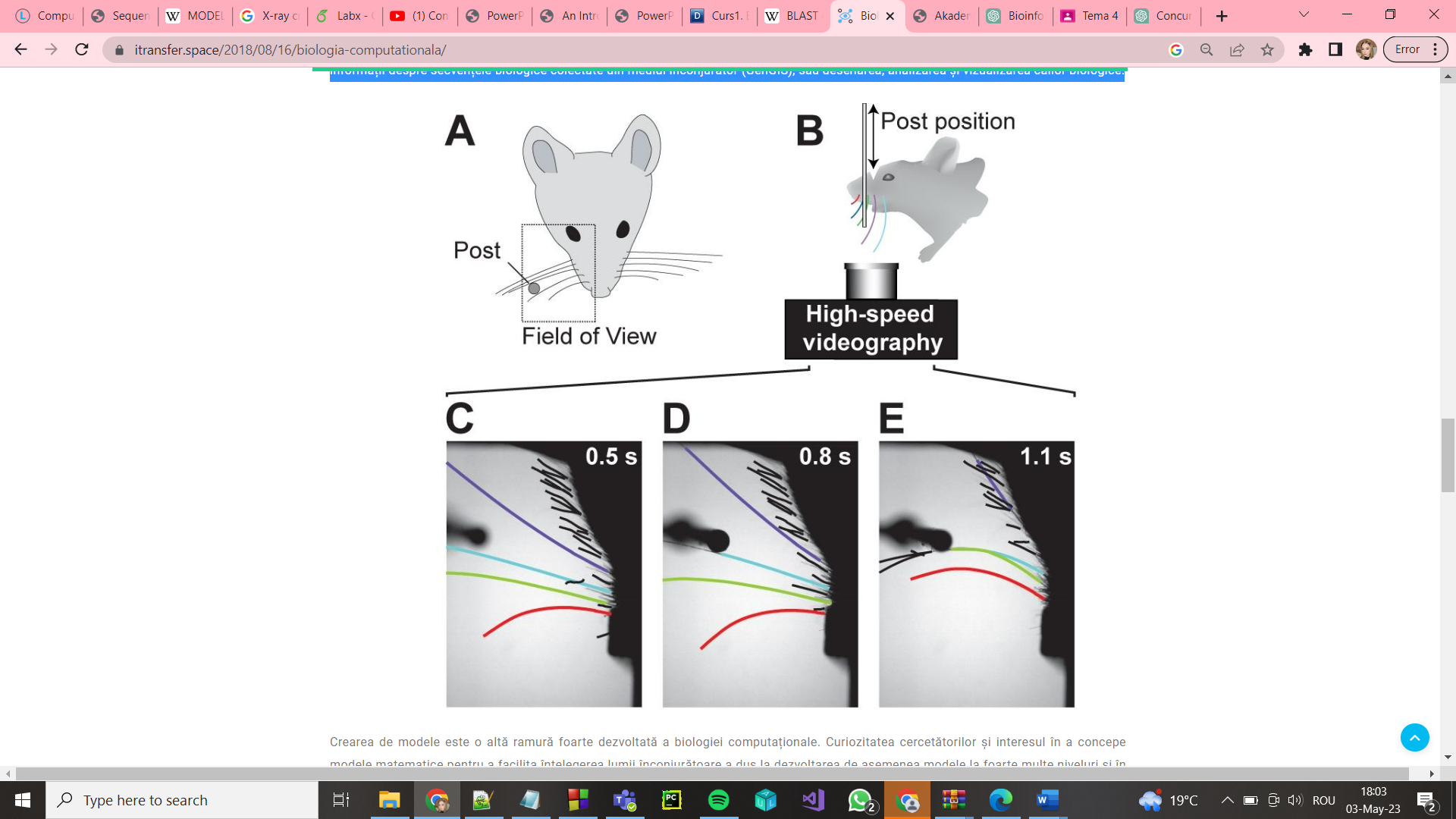
Bioinformatica și **informatica computațională** se intersectează în dezvoltarea de algoritmi și metode eficiente pentru analiza și interpretarea datelor biologice. Informatica computațională aduce concepte și tehnici de programare, algoritmi, baze de date și sisteme de calcul pentru a rezolva probleme specifice din domeniul bioinformaticii. Colaborarea dintre cele două subdomenii contribuie la progresul în domeniul analizei datelor biologice.

Există trei componenente principale ale bioinformaticii: analiza datelor, dezvoltarea de software-uri și modelarea datelor.

Partea de analiza de date are ca scop procesarea unor date neprelucrate în vederea obținerii unui mod mai facil de înțelegere a acestora. Aici poate fi vorba de aspecte de **statistică, data mining sau doar crearea de interpretări vizuale**. Practic, prin acest fel se ajunge la o modalitate mai simplă de a interpreta informația, ceea ce o face mai accesibilă și mai ușor de transmis. Datele procesate corepunzător pot fi raportate, împărtășite sau explicate unui public mai larg. Acest lucru este foarte util pentru diseminarea de informații și promovarea de colaborări între arii de studiu similare. Dat fiind că biologia, sau științele vieții în general, implică o înțelegere foarte vastă și utilizarea de informații din domenii similare dar foarte diferite în același timp, facilitarea transferului de date prin abordări computaționale este foarte utilă.

Dezvoltarea de softwareuri pentru explorarea științelor biologice este o direcție nouă de dezvoltare a cercetării din lumea informaticii, și este din ce în ce mai populară. Softwareurile open source oferă o platformă ideală pentru dezvoltarea aplicațiilor de biologie computațională. Open source presupune accesul mai multor persoane/entitați care pot contribui sau beneficia de softwareul respectiv. Sunt patru calități care fac sistemele open source foarte potrivite pentru acest scop. Reproductibilitatea acestora permite cercetătorilor să folosească aceleași metode exacte folosite pentru a calcula relațiile dintre datele biologice. Dezvoltarea accelerata se referă la faptul că dezvoltatorii și cercetătorii nu trebuie să reinventeze algoritmii deja existenți pentru taskuri minore. În schimb, pot folosi programe deja existente pentru a salva timp pentru dezavoltarea și implementarea unor proiecte mai mari. Calitatea crescută a rezultatelor provenită de la faptul ca mai mulți experți din același domeniu pot contribui la crearea software-ului aduce o siguranță în plus și scade predispoziția codului la erori. Nu în ultimul rând, software-urile open source au o valabilitate pe termen mult mai lung. Aceste programe nu sunt asociate nici unei întreprinderi și nu se află sub vreun brevet, ceea ce le permite distribuirea pe mai multe pagini de internet și asigură valabilitatea lor în viitor.

Software-urile care pot fi create în domeniul biologiei computaționale sunt atât de variate precum e lumea naturală. Exemplele includ urmărirea mustăților rozătoarelor, recunoșterea (semi-)automată a diverselor grupuri taxonomice, combinarea de date cartografice cu informații despre secvențele biologice colectate din mediul înconjurător, sau desenarea, analizarea și vizualizarea căilor biologice.



Creșterea simulată a plantelor este o altă provocare importantă a biologiei sistemelor și a biologiei matematice, care urmărește reproducerea morfologiei plantelor cu ajutorul simulărilor computaționale. Arborii electronici (e-arbori) utilizează de obicei sisteme Lindenmayer pentru a simula creșterea. Sistemele L sunt foarte importante în domeniul științei complexității și al vieții artificiale. O potențială problemă pentru simularea creșterii plantelor este aceea de a integra în mod constant factorii de mediu, cum ar fi plantele înconjurătoare, obstacolele, disponibilitatea apei și mineralelor și condițiile de iluminare. Pentru acest lucru trebuie construite medii virtuale cu un număr parametri fezabili din punct de vedere computațional, ceea ce nu numai că simulează creșterea plantei, ci și mediul în care se dezvoltă. Schimbările în disponibilitatea resurselor influențează creșterea plantelor, ceea ce are ca rezultat o nouă schimbare a disponibilității resurselor. Modele din ce n ce mai puternice și robuste vor fi necesare pentru a simula eficient aceste interacțiuni recursive.

**Persoane importante:**

Eugene Myers - este un cercetător și profesor de bioinformatică la Universitatea din California, Berkeley. A contribuit la dezvoltarea algoritmului BLAST și a fost un pionier în domeniul secvențierii ADN.

Temple F. Smith - este unul dintre pionierii bioinformaticii și fondatorul domeniului aliniamentului secvențelor. A contribuit la dezvoltarea algoritmului Smith-Waterman, un algoritm esențial în alinierea secvențelor biologice.

5. Steven Salzberg - este un profesor și cercetător în bioinformatică la Universitatea Johns Hopkins. A fost implicat în dezvoltarea unor instrumente cheie în secvențierea genomului și analiza datelor genomice.

Există numeroase foruri importante în domeniul bioinformaticii, care includ reviste și conferințe recunoscute pe plan internațional. Iată câteva exemple:

**Reviste:**

Bioinformatics: Este o revistă de referință în bioinformatică, care publică articole originale de cercetare, metode, aplicații și perspective în domeniu.

Nucleic Acids Research: Această revistă publică articole în domeniul bioinformaticii, inclusiv baze de date și instrumente software utilizate în analiza datelor biologice.

PLOS Computational Biology: Este o revistă deschisă de acces, care se concentrează pe intersecția dintre biologie și informatică, inclusiv bioinformatică și modelare computațională.

**Conferințe:**

International Conference on Intelligent Systems for Molecular Biology (ISMB): Este una dintre cele mai prestigioase conferințe în domeniul bioinformaticii, care acoperă o gamă largă de subiecte, inclusiv analiza secvențelor, modelarea moleculară și medicina personalizată.

ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB): Este o conferință care aduce împreună cercetători și profesioniști din domeniul bioinformaticii, biologiei computaționale și informaticii medicale.

Intelligent Systems for Molecular Biology (ISMB)/European Conference on Computational Biology (ECCB): Este o conferință care se desfășoară anual și reunește comunitatea internațională în domeniul bioinformaticii și biologiei computaționale.

Bibliografie:

Bioinformatică / Gheorghe-Ioan Mihalaş, Anca Tudor, Sorin Paralescu. - Timişoara: Editura Victor Babeş, 2011

Daniela Zaharie – curs 5 - Tehnici algoritmice utilizate in bioinformatica