Cover Sheet: Request 16075

CIS 4XXX - Introduction to Bioinformatics

Info

Process	Course|New|Ugrad/Pro
Status	Pending at PV - University Curriculum Committee (UCC)
Submitter	Christina Boucher christinaboucher@ufl.edu
Created	4/8/2021 12:33:15 PM
Updated	5/13/2021 4:38:05 PM
Description of request	A request for instatement of an undergraduate class in CISE entitled "Introduction to Bioinformatics"

Actions

Step	Status	Group	User	Comment	Updated
Department	Approved	ENG - Computer and Information Science and Engineering 19140000	Christina Gardner-McCune		4/23/2021
No document changes					
College	Approved	ENG - College of Engineering	Heidi Dublin	Approved by HWCOE Curriculum Committee and Faculty Council	5/13/2021
EXAMPLE-ASSIGNMENTS.docx BioinformaticsCISE-course-approval.pdf BioinformaticsCISE-course-conflict2.pdf HWCOE-Undergrad-Course-Syllabus-F2F-Fall-2020-introduction-to-bioinformatics.docx					5/3/2021 5/5/2021 5/5/2021 5/5/2021
University Curriculum Committee	Pending	PV - University Curriculum Committee (UCC)			5/13/2021
No document changes					
Statewide Course Numbering System					
No document changes					
Office of the Registrar					
No document changes					
Catalog					
No document changes					
Student Academic Support System					
No document changes					
College Notified					
No document changes					

Course|New for request 16075

Info

Request: CIS 4XXX - Introduction to Bioinformatics
Description of request: A request for instatement of an undergraduate class in CISE entitled "Introduction to Bioinformatics"
Submitter: Christina Boucher christinaboucher@ufl.edu
Created: 9/13/2021 10:09:46 AM
Form version: 5

Responses

Recommended Prefix
Enter the three letter code indicating placement of course within the discipline (e.g., POS, ATR, ENC). Note that for new course proposals, the State Common Numbering System (SCNS) may assign a different prefix.

Response:
CIS

Course Level
*Select the one digit code preceding the course number that indicates the course level at which the course is taught (e.g., 1=freshman, 2=sophomore, etc.). *
Note: 5000 level courses must be submitted through the undergraduate new course process

Response:
4

Course Number
Enter the three digit code indicating the specific content of the course based on the SCNS taxonomy and course equivalency profiles. For new course requests, this may be XXX until SCNS assigns an appropriate number.

Response:
XXX

Category of Instruction
*Indicate whether the course is introductory, intermediate or advanced. Introductory courses are those that require no prerequisites and are general in nature. Intermediate courses require some prior preparation in a related area. Advanced courses require specific competencies or knowledge relevant to the topic prior to enrollment. *

Response:
Intermediate

• 1000 level = Introductory undergraduate
• 2000 level = Introductory undergraduate
• 3000 level = Intermediate undergraduate
• 4000 level = Advanced undergraduate
• 5000 level = Introductory graduate
• 6000 level = Intermediate graduate
• 7000 level = Advanced graduate
• 4000/5000= Joint undergraduate/graduate
• 4000/6000= Joint undergraduate/graduate

Joint undergraduate/graduate courses must be approved by the UCC and the Graduate Council)

Lab Code
Enter the lab code to indicate whether the course is lecture only (None), lab only (L), or a combined lecture and lab (C).

Response:
None

Course Title
*Enter the title of the course as it should appear in the Academic Catalog. There is a 100 character limit for course titles. *

Response:
Introduction to Bioinformatics

Transcript Title
Enter the title that will appear in the transcript and the schedule of courses. Note that this must be limited to 30 characters (including spaces and punctuation).

Response:
Introduction to Bioinformatics

Degree Type
Select the type of degree program for which this course is intended.

Response:
Baccalaureate

Delivery Method(s)
Indicate all platforms through which the course is currently planned to be delivered.

Response:
On-Campus

Co-Listing
Will this course be jointly taught to undergraduate, graduate, and/or professional students?

Response:
No

Effective Term
Select the requested term that the course will first be offered. Selecting "Earliest" will allow the course to be active in the earliest term after SCNS approval. If a specific term and year are selected, this should reflect the department's best projection. Courses cannot be implemented retroactively, and therefore the actual effective

term cannot be prior to SCNS approval, which must be obtained prior to the first day of classes for the effective term. SCNS approval typically requires 2 to 6 weeks after approval of the course at UF.

Response:
Fall

Effective Year
Select the requested year that the course will first be offered. See preceding item for further information.

Response:
2022

Rotating Topic?
Select "Yes" if the course can have rotating (varying) topics. These course titles can vary by topic in the Schedule of Courses.

Response:
Yes

Repeatable Credit?
Select "Yes" if the course may be repeated for credit. If the course will also have rotating topics, be sure to indicate this in the question above.

Response:
No

Amount of Credit
Select the number of credits awarded to the student upon successful completion, or select "Variable" if the course will be offered with variable credit and then indicate the minimum and maximum credits per section. Note that credit hours are regulated by Rule 6A-10.033, FAC. If you select "Variable" for the amount of credit, additional fields will appear in which to indicate the minimum and maximum number of total credits.

Response:
3

S/U Only?
Select "Yes" if all students should be graded as S/U in the course. Note that each course must be entered into the UF curriculum inventory as either letter-graded or S/U. A course may not have both options. However, letter-graded courses allow students to take the course S/U with instructor permission.

Response:
No

Contact Type
Select the best option to describe course contact type. This selection determines whether base hours or headcount hours will be used to determine the total contact hours per credit hour. Note that the headcount hour options are for courses that involve contact between the student and the professor on an individual basis.

Response:
Regularly Scheduled

• *Regularly Scheduled [base hr]*
• *Thesis/Dissertation Supervision [1.0 headcount hr]*
• *Directed Individual Studies [0.5 headcount hr]*
• *Supervision of Student Interns [0.8 headcount hr]*
• *Supervision of Teaching/Research [0.5 headcount hr]*
• *Supervision of Cooperative Education [0.8 headcount hr]*

Contact the Office of Institutional Planning and Research (352-392-0456) with questions regarding contact type.

Weekly Contact Hours
Indicate the number of hours instructors will have contact with students each week on average throughout the duration of the course.

Response:
3

Course Description
Provide a brief narrative description of the course content. This description will be published in the Academic Catalog and is limited to 500 characters or less. See course description guidelines.

Response:
This undergraduate course follows Compeau and Pevzner's active learning approach to learning bioinformatics. Following their "Bioinformatics Algorithms" textbook, it covers fundamental topics in genomics, machine learning, and biological data science, it strikes a unique balance between practical challenges in modern biology and fundamental algorithmic ideas, thus capturing the interest of students of both biology and computer science. ,

Prerequisites
Indicate all requirements that must be satisfied prior to enrollment in the course. Prerequisites will be automatically checked for each student attempting to register for the course. The prerequisite will be published in the Academic Catalog and must be formulated so that it can be enforced in the registration system. Please note that upper division courses (i.e., intermediate or advanced level of instruction) must have proper prerequisites to target the appropriate audience for the course.
Courses level 3000 and above must have a prerequisite.
Please verify that any prerequisite courses listed are active courses.

Response:
COT 3100 (C) and COP 3503 (C)

Completing Prerequisites on UCC forms:

• *Use "&" and "or" to conjoin multiple requirements; do not used commas, semicolons, etc.*
• *Use parentheses to specify groupings in multiple requirements.*
• *Specifying a course prerequisite (without specifying a grade) assumes the required passing grade is D-. In order to specify a different grade, include the grade in parentheses immediately after the course number. For example, "MAC 2311(B)" indicates that students are required to obtain a grade of B in Calculus I. MAC2311 by itself would only require a grade of D-.*
• *Specify all majors or minors included (if all majors in a college are acceptable the college code is sufficient).*
• *"Permission of department" is always an option so it should not be included in any prerequisite or co-requisite.*
• *If the course prerequisite should list a specific major and/or minor, please provide the plan code for that major/minor (e.g., undergraduate Chemistry major = CHY_BS, undergraduate Disabilities in Society minor =*

DIS_UMN)

Example: A grade of C in HSC 3502, passing grades in HSC 3057 or HSC 4558, and undergraduate PBH student should be written as follows: HSC 3502(C) & (HSC 3057 or HSC 4558) & UGPBH
* *

Co-requisites
Indicate all requirements that must be taken concurrently with the course. Co-requisites are not checked by the registration system. If there are none please enter N/A.

Response:
None.

Rationale and Placement in Curriculum
Explain the rationale for offering the course and its place in the curriculum.

Response:
Large biological datasets pose interesting challenges in computer science, as they require specialized algorithms that are also scalable with respect to theoretical bounds. In this course, several algorithmic paradigms will be considered and implemented by the students, allowing them to gain experience in analyzing biological data, understanding and formulating data analysis questions, and implementing in a manner that it remains scalable. There is a growing interest in data science and bioinformatics industry as many new companies are created, and different divisions in existing companies are being developed, such as Google Health. There is no course in bioinformatics for undergraduate students in CISE. This will give the students the experience needed to pursue bioinformatics in their academic or industry endeavors.

Course Objectives
Describe the core knowledge and skills that student should derive from the course. The objectives should be both observable and measurable.

Response:
Modern biological techniques are generating a variety of large-scale data, and require sophisticated algorithms for their analysis. Tasks include assembling the genomic sequence, predicting the protein-coding regions, their function and the manner in which they are regulated. The course will provide a broad overview of the computational techniques currently used in bioinformatics. Students completing the course will be able to analyze a variety of biological data, a skill they will develop through meetings with biological partners and groups projects, which will consist of algorithm implementation and application. More specifically, students will learn and directly apply:
• Dynamic programming for sequence alignment,
• Hidden Markov models for gene finding, and
• Graph algorithms for sequence assembly and motif finding.
In this course, students interact with the material through reading relevant literature, participating in group discussions, creating relevant presentations, working on an individual or group project, and listening to guest speakers. Students will apply their knowledge in assignments where they design, implement, and evaluate a bioinformatics systems.

Course Textbook(s) and/or Other Assigned Reading
*Enter the title, author(s) and publication date of textbooks and/or readings that will be assigned. Please provide specific examples to evaluate the course and identify required textbooks. *

Response:

The following textbook is recommended but not required: Bioinformatics Algorithms: An active learning approach. 3rd edition. By Phillip Compeau and Pavel Pevzner. 2014.

Weekly Schedule of Topics
Provide a projected weekly schedule of topics. This should have sufficient detail to evaluate how the course would meet current curricular needs and the extent to which it overlaps with existing courses at UF.

Response:
Week 1 : Introduction
Weeks 2 and 3: Algorithm basics. Review of intractability, running time, and basic data structures
Week 4 : Biology basics. DNA, RNA and Protein. Central Dogma of Biology.
Week 5 : Genome Sequencing
Week 6 : Sequence Alignment: Motivation and Dynamic programming algorithm
Week 7 : Greedy algorithms for solving motif finding.
Weeks 8 and 9 : Graph algorithms and models for genome assembly.
Week 10 : How do we sequencing antibiotics? Brute Force Algorithms
Week 11 : Succinct data structures for searching optical mapping data.
Week 12 : Are there fragile regions in the genome? Combinatorics.
Weeks 13 and 14 : How did yeast become a wine maker? Machine learning.
Week 15: Class presentations
Week 15: Final examination

Grading Scheme
List the types of assessments, assignments and other activities that will be used to determine the course grade, and the percentage contribution from each. This list should have sufficient detail to evaluate the course rigor and grade integrity. Include details about the grading rubric and percentage breakdowns for determining grades. If participation and/or attendance are part of the students grade, please provide a rubric or details regarding how those items will be assessed.

Response:
Evaluation of Grades:
Assignment | Total Points | Percentage of Final Grade
Homework Sets (3) | 100 each | 90%
Class participation | 100

| 10%

Grading Rubric:
Percent | Grade | Grade Points
94 - 100 | A | 4.00
90 - 93 | A- | 3.67
87 - 89 | B+ | 3.33
84 - 86 | B | 3.00
80 - 83 | B- | 2.67
75 - 79 | C+ | 2.33
70 - 74 | C | 2.00
60 - 69 | D | 1.00
0 - 59 | E | 0.00

Instructor(s)
Enter the name of the planned instructor or instructors, or "to be determined" if instructors are not yet identified.

Response:

Christina Boucher
Associate Professor
Computer & Information Science & Engineering Department
Herbert Wertheim College of Engineering
University of Florida
Gainesville, FL 32611
http://www.christinaboucher.com/
Google Scholar

and "to be determined"

Attendance & Make-up
Please confirm that you have read and understand the University of Florida Attendance policy.
A required statement statement related to class attendance, make-up exams and other work will be included in the syllabus and adhered to in the course. Courses may not have any policies which conflict with the University of Florida policy. The following statement may be used directly in the syllabus.

• *Requirements for class attendance and make-up exams, assignments, and other work in this course are consistent with university policies that can be found at:*
https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx

Response:
Yes

Accomodations
Please confirm that you have read and understand the University of Florida Accommodations policy.
A statement related to accommodations for students with disabilities will be included in the syllabus and adhered to in the course. The following statement may be used directly in the syllabus:

• *Students with disabilities requesting accommodations should first register with the Disability Resource Center (352-392-8565, www.dso.ufl.edu/drc/) by providing appropriate documentation. Once registered, students will receive an accommodation letter which must be presented to the instructor when requesting accommodation. Students with disabilities should follow this procedure as early as possible in the semester.*

Response:
Yes

UF Grading Policies for assigning Grade Points
Please confirm that you have read and understand the University of Florida Grading policies.
Information on current UF grading policies for assigning grade points is require to be included in the course syllabus. The following link may be used directly in the syllabus:

• *https://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx*

Response:
Yes

Course Evaluation Policy
Course Evaluation Policy
Please confirm that you have read and understand the University of Florida Course Evaluation Policy.
A statement related to course evaluations will be included in the syllabus. The following statement may be used

directly in the syllabus:

• Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at https://gatorevals.aa.ufl.edu/public-results/. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via https://ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at https://gatorevals.aa.ufl.edu/public-results/.

* *

Response:
Yes

Introduction to Bioinformatics
Class Periods: TBD
Location: TBD
Academic Term: Fall 2020

Instructor:
Christina Boucher
christinaboucher@ufl.edu
CSE 546
Office Hours: TBD

Teaching Assistant/Peer Mentor/Supervised Teaching Student:
Please contact through the Canvas website
- TBD

Course Description
Introduces essential topics bioinformatics and analysis of biological data. Focus is on the algorithms that are behind bioinformatic tools. It covers fundamental topics in genomics, machine learning, and combinatorics. Each topic begins with a question or problem related to biology or science and subsequently introduces and an algorithmic or combinatorial approach to solving it. Each topic is complemented by an assignment in which the students implement and apply algorithms learned in class to analyze real biological data.

Course Pre-Requisites / Co-Requisites
A grade of C or higher in both COT 3503 and COT 3100.

Course Objectives
Modern biological techniques are generating a variety of large-scale data, and require sophisticated algorithms for their analysis. Tasks include assembling the genomic sequence, predicting the protein-coding regions, their function and the manner in which they are regulated. The course will provide a broad overview of the computational techniques currently used in bioinformatics. Students completing the course will be able to analyze a variety of biological data, a skill they will develop through meetings with biological partners and groups projects, which will consist of algorithm implementation and application. More specifically, students will <u>learn</u> and <u>directly apply</u>:

- Dynamic programming for sequence alignment,
- Hidden Markov models for gene finding, and
- Graph algorithms for sequence assembly and motif finding.

In this course, students interact with the material through reading relevant literature, participating in group discussions, creating relevant presentations, working on an individual or group project, and listening to guest speakers. Students will apply their knowledge in a research project where they design, implement, and evaluate a bioinformatics system.

Materials and Supply Fees
None.

Required Textbooks and Software
None.

Recommended Materials
Bioinformatics Algorithms: An active learning approach. 3rd edition. By Phillip Compeau and Pavel Pevzner. 2014.

Course Schedule
Week 1:   Introduction / Course Slides

Weeks 2 and 3:   Algorithm basics. Review of intractability, running time, and basic data structures / Course Slides

Week 4:   Biology basics. DNA, RNA and Protein. Central Dogma of Biology / Course Slides

Week 5:   Genome Sequencing / Course Slides / Chapter 3 of Bioinformatics Algorithms / *Assignment 1 posted (Implementation and analysis of sequence alignment files)*

Week 6:   Sequence Alignment: Motivation and Dynamic programming algorithm / Course Slides / Chapter 4 of Bioinformatics Algorithms

Week 7:   Greedy algorithms for solving motif finding / Course Slides / Chapter 2 of Bioinformatics Algorithms / *Assignment 1 due / Assignment 2 posted (Implementation of algorithms to align two sequences using dynamic programming)*

Weeks 8 and 9:   Graph algorithms and models for genome assembly / Course Slides

Week 10:   How do we sequencing antibiotics? Brute Force Algorithms / Course Slides / Chapter 4 of Bioinformatics Algorithms / *Assignment 2 due / Assignment 3 posted (Implementation of genome assembly algorithms)*

Week 11:   Succinct data structures for searching optical mapping data / Course Slides

Week 12:   Are there fragile regions in the genome? Combinatorics / Course Slides / Chapter 6 of Bioinformatics Algorithms / *Assignment 3 due*

Weeks 13 and 14 :   How did yeast become a wine maker? Machine learning / Course Slides / Chapter 7 of Bioinformatics Algorithms

Weeks 15:   Class presentations

F2F Course Policy in Response to COVID-19

We will have face-to-face instructional sessions to accomplish the student learning objectives of this course. In response to COVID-19, the following policies and requirements are in place to maintain your learning environment and to enhance the safety of our in-classroom interactions.

- You are required to wear approved face coverings at all times during class and within buildings. Following and enforcing these policies and requirements are all of our responsibility. Failure to do so will lead to a report to the Office of Student Conduct and Conflict Resolution.

- This course has been assigned a physical classroom with enough capacity to maintain physical distancing (6 feet between individuals) requirements. Please utilize designated seats and maintain appropriate spacing between students. Please do not move desks or stations.

- Sanitizing supplies are available in the classroom if you wish to wipe down your desks prior to sitting down and at the end of the class.

- Follow your instructor's guidance on how to enter and exit the classroom. Practice physical distancing to the extent possible when entering and exiting the classroom.

- If you are experiencing COVID-19 symptoms (Click here for guidance from the CDC on symptoms of coronavirus), please use the UF Health screening system and follow the instructions on whether you are able to attend class. Click here for UF Health guidance on what to do if you have been exposed to or are experiencing Covid-19 symptoms.
- Course materials will be provided to you with an excused absence, and you will be given a reasonable amount of time to make up work. Find more information in the university attendance policies.

Attendance Policy, Class Expectations, and Make-Up Policy

 ]Attendance will be not taken but there will be periodic activities and discussions throughout the semester that students should attend most of. Cell phones and laptops are allowed but please put them on mute during class. 10% will be deducted for every 24 hours that a homework is late. After 48 hours the homework will receive a zero. If the class size continues to be small, the deadline is negotiable. Try to talk to me if you need an extensions. Extensions will be applied to everyone.   All suspected plagiarism and cheating cases will be evaluated by the SCCR. This is consistent with University of Florida policy. A grade of zero will be given to all cases of plagiarism and cheating. Excused absences must be in compliance with university policies in the Graduate Catalog (http://gradcatalog.ufl.edu/content.php?catoid=10&navoid=2020#attendance) and require appropriate documentation.

Evaluation of Grades

Assignment	Total Points	Percentage of Final Grade
Homework Sets (3)	100 each	90%
Class Presentation	100	10%
		100%

Grading Policy

Percent	Grade	Grade Points
93.4 - 100	A	4.00
90.0 - 93.3	A-	3.67
86.7 - 89.9	B+	3.33
83.4 - 86.6	B	3.00
80.0 - 83.3	B-	2.67
76.7 - 79.9	C+	2.33

73.4 - 76.6	C	2.00
70.0 - 73.3	C-	1.67
66.7 - 69.9	D+	1.33
63.4 - 66.6	D	1.00
60.0 - 63.3	D-	0.67
0 - 59.9	E	0.00

More information on UF grading policy may be found at:
https://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx

Students Requiring Accommodations
Students with disabilities who experience learning barriers and would like to request academic accommodations should connect with the disability Resource Center by visiting https://disability.ufl.edu/students/get-started/. It is important for students to share their accommodation letter with their instructor and discuss their access needs, as early as possible in the semester.

Course Evaluation
Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at https://gatorevals.aa.ufl.edu/students/. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via https://ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at https://gatorevals.aa.ufl.edu/public-results/.

University Honesty Policy
UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code (https://sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Commitment to a Safe and Inclusive Learning Environment
The Herbert Wertheim College of Engineering values broad diversity within our community and is committed to individual and group empowerment, inclusion, and the elimination of discrimination. It is expected that every person in this class will treat one another with dignity and respect regardless of gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture.

If you feel like your performance in class is being impacted by discrimination or harassment of any kind, please contact your instructor or any of the following:
• Your academic advisor or Graduate Program Coordinator
• Robin Bielling, Director of Human Resources, 352-392-0903, rbielling@eng.ufl.edu
• Curtis Taylor, Associate Dean of Student Affairs, 352-392-2177, taylor@eng.ufl.edu
• Toshikazu Nishida, Associate Dean of Academic Affairs, 352-392-0943, nishida@eng.ufl.edu

Software Use
All faculty, staff, and students of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. We, the members of the University of Florida community, pledge to uphold ourselves and our peers to the highest standards of honesty and integrity.

Student Privacy

There are federal laws protecting your privacy with regards to grades earned in courses and on individual assignments. For more information, please see: https://registrar.ufl.edu/ferpa.html

Campus Resources:

<u>Health and Wellness</u>

U Matter, We Care:
Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact umatter@ufl.edu so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at 352-392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.

Counseling and Wellness Center: http://www.counseling.ufl.edu/cwc, and 392-1575; and the University Police Department: 392-1111 or 9-1-1 for emergencies.

Sexual Discrimination, Harassment, Assault, or Violence
If you or a friend has been subjected to sexual discrimination, sexual harassment, sexual assault, or violence contact the **Office of Title IX Compliance**, located at Yon Hall Room 427, 1908 Stadium Road, (352) 273-1094, title-ix@ufl.edu

Sexual Assault Recovery Services (SARS)
Student Health Care Center, 392-1161.

University Police Department at 392-1111 (or 9-1-1 for emergencies), or http://www.police.ufl.edu/.

<u>Academic Resources</u>

E-learning technical suppor*t*, 352-392-4357 (select option 2) or e-mail to Learning-support@ufl.edu. https://lss.at.ufl.edu/help.shtml.

Career Resource Center, Reitz Union, 392-1601. Career assistance and counseling. https://www.crc.ufl.edu/.

Library Support, http://cms.uflib.ufl.edu/ask. Various ways to receive assistance with respect to using the libraries or finding resources.

Teaching Center, Broward Hall, 392-2010 or 392-6420. General study skills and tutoring. https://teachingcenter.ufl.edu/.

Writing Studio, 302 Tigert Hall*, 846-1138. Help brainstorming, formatting, and writing papers. https://writing.ufl.edu/writing-studio/.

Student Complaints Campus*: https://care.dso.ufl.edu.

On-Line Students Complaints*: http://www.distance.ufl.edu/student-complaint-process.

From: **Zajac-Kaye,Maria** mzajackaye@ufl.edu
Subject: Re: Bioinformatics@CISE
Date: October 10, 2020 at 4:41 PM
To: Boucher,Christina A cboucher@cise.ufl.edu
Cc: Zakshevsky, Mark J zmz30@ufl.edu

Hi Christina,

Sorry for the late response. I am happy to help and sign the consultation form.
Could you please write the comments on the consultation form and send to my assistant (cc here) so he could help with details of the form.
Thanks
Maria
--
Maria Zajac-Kaye, PhD
Professor and Interim Chair
Department of Anatomy and Cell Biology
co-Director Cancer Biology Concentration
University of Florida
Cancer and Genetics Research Complex
2033 Mowry Rd. Rm. 360
Gainesville, Florida 32610
Tel: 352 273-9153
email: mzajackaye@ufl.edu

From: "Boucher,Christina A" <cboucher@cise.ufl.edu>
Date: Tuesday, October 6, 2020 at 3:36 PM
To: "Zajac-Kaye,Maria" <mzajackaye@ufl.edu>
Subject: Bioinformatics@CISE

Dear Maria,

I am a faculty member in CISE. My research area is in bioinformatics. I have been teaching an undergraduate class in bioinformatics in the last three years as a special topics class. Attached is the syllabus. I am currently applying for it to be an official class in the CISE dept. Although this should not conflict with any of your classes (it is about the algorithms behind the bioinformatics software), I was wondering if you could fill out an external consultation form for me.
 See: https://approval.ufl.edu/policies/external-consultations/. It says it must be filled out by the dept chair.

Let me know if you would discuss this further. I am happy to via email or zoom.

Best,
Christina

Christina Boucher
Associate Professor
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, FL
christinaboucher@ufl.edu
www.christinaboucher.com
Christina Boucher@Google Scholar

From: **Wayne,Marta L** mlwayne@ufl.edu
Subject: Re: Bioinformatics@CISE
Date: October 6, 2020 at 7:38 PM
To: Boucher,Christina A cboucher@cise.ufl.edu

MW

Thanks— I am not hearing about any conflicts. :)

Get Outlook for iOS

From: Boucher,Christina A <cboucher@cise.ufl.edu>
Sent: Tuesday, October 6, 2020 7:14:49 PM
To: Boucher,Christina A <cboucher@cise.ufl.edu>
Cc: Wayne,Marta L <mlwayne@ufl.edu>; Braun,Edward Louis <ebraun68@ufl.edu>
Subject: Re: Bioinformatics@CISE

Actually, I clicked on the link and it is still live. If you go to the "Progress" tab at the top, you will the list of topics that we cover and the order. Here is the link:
https://www.christinaboucher.com/teaching/4930/home_progress.php

I can add you to Canvas if you would like the course slides.   This semester I am also videotaping all lectures so they are available on Canvas too.

Best,
Christina

Christina Boucher
Associate Professor
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, FL
christinaboucher@ufl.edu
www.christinaboucher.com
Christina Boucher@Google Scholar

On Oct 6, 2020, at 6:05 PM, Boucher,Christina A <cboucher@cise.ufl.edu> wrote:

Thanks. I recently took down the link because I found that someone had stollen the course from another university. I now just use Canvas. I will the link back up temporarily and email it to you. But yes, we use the new Pevzner and Compeau textbook and it should not really conflict because we discuss/implement the algorithms behind some of the tools. For example, the first assignment is to implement Smith Waterman and show how it scales w.r.t. sequence length.

I hope that helps!

Best,
Christina

Christina Boucher
Associate Professor
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, FL
christinaboucher@ufl.edu
www.christinaboucher.com
Christina Boucher@Google Scholar

On Oct 6, 2020, at 3:52 PM, Wayne,Marta L <mlwayne@ufl.edu> wrote:

Dear Christina,

Please see below.

Thanks!
Marta

Get Outlook for iOS

From: Braun,Edward Louis <ebraun68@ufl.edu>
Sent: Tuesday, October 6, 2020 3:48 PM
To: Wayne,Marta L; Burleigh, Gordon; Gitzendanner, Matt; Ponciano Castellanos,Jose Miguel; Barbazuk,William Bradley
Subject: Re: Bioinformatics@CISE

Marta,

I'm not seeing an actual schedule of topics on the syllabus. She does link to a web page with topics (https://www.christinaboucher.com/teaching/4930/index.php), but all it says is that the topics will include:

Dynamic programming for sequence alignment,
Hidden Markov models for gene finding, and
Graph algorithms for sequence assembly and motif finding

Graph algorithms for sequence assembly and motif finding.

I'm familiar with the Pevzner and Compeau textbook she is following, and I doubt that anything in it overlaps with things we'd want to teach (if I recall correctly Pevzner and Compeau it is really a pretty hardcore "here is how Smith-Waterman works" kind of book - but take that statement with the caveat that my copy is in my office and not home where I can check right now)

I don't want to be a stickler and demand a precise schedule when classes like this tend to change, but it would be nice to have something more detailed than what I copied above. She'll probably need that for UCC anyway

Ed

From: Wayne,Marta L <mlwayne@ufl.edu>
Sent: Tuesday, October 6, 2020 3:12 PM
To: Burleigh, Gordon <gburleigh@ufl.edu>; Gitzendanner, Matt <magitz@ufl.edu>; Ponciano Castellanos,Jose Miguel <josemi@ufl.edu>; Braun,Edward Louis <ebraun68@ufl.edu>; Barbazuk,William Bradley <bbarbazuk@ufl.edu>
Subject: FW: Bioinformatics@CISE

From: "Boucher,Christina A" <cboucher@cise.ufl.edu>
Date: Tuesday, October 6, 2020 at 2:55 PM
To: Marta Wayne <mlwayne@ufl.edu>
Subject: Bioinformatics@CISE

Dear Marta,

I am a faculty member in CISE. My research area is in bioinformatics. I have been teaching an undergraduate class in bioinformatics in the last three years as a special topics class. Attached is the syllabus. I am currently applying for it to be an official class in the CISE dept. Although this should not conflict with any of your classes (it is about the algorithms behind the bioinformatics software), I was wondering if you could fill out an external consultation form for me. See: https://approval.ufl.edu/policies/external-consultations/. It says it must be filled out by the dept chair.

Let me know if you would discuss this further. I am happy to via email or zoom.

Best,
Christina

Christina Boucher
Associate Professor
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, FL
christinaboucher@ufl.edu
www.christinaboucher.com
Christina Boucher@Google Scholar

ASSIGNMENT 1: LOOKING AT ALIGNMENT FURTHER

Important notes:

- Your programs should be written in any programming language.
- You do not have to use linux/unix for this assignment but I recommend it
- You can choose to this assignment in partners or alone.
- The name of your program should be [your-lastname1-lastname2].{c,cpp,java}, e.g. boucher-ruiz.cpp. If you would like to attach a readme text file, it should be named [your-lastname1-lastname2]-readme.txt
- For (e), please hand-in a .docx or .pdf file.
- Put everything in a .zip and submit it via Canvas.
- As before, please give me any comments on how you found the difficulty of the assignment. It helps me prepare the next assignment.

Problem Statement: Given two DNA sequences, we are interested in finding the best pairwise sequence alignment between them. The scoring function uses 2 for any match, -1 for any mismatch, and -2 for any match against a gap.

More specifically, we are interested in the best alignment, its corresponding alignment score, the possibility of existing multiple optimal alignments, and, if the answer is positive, the set of all optimal alignments. You can assume the two DNA sequences will be given to you in a fasta file that will be specified on the command line. Note that each part of the problem must be written in a separate output file (this simplifies the grading).

(a) (output: assignment1.o1): 20 points
In this file, write the optimal score as an integer value.

(b) (output: assignment1.o2): 20 points
For this part, you should write the dynamic programming matrix.
Assume the first DNA has length m and the second one has length n. Your output, assignment1.o2, should contain m+1 lines each containing n+1 integer values separated by a single space character. The value j in the line i corresponds the value in column j and row i of the dynamic programming matrix.

(c) (output: assignment1.o3): 20 points
For this part, we are interested in one optimal alignment. Write the best alignment of the two input sequences on two lines. For any gap write a single character '-'. Notice that there might be multiple optimal alignments and you are required to only report one.

(d) (output: assignment1.o4): 20 points
Are there multiple optimal alignments? Write either 'YES' or 'NO' (all in capital letters) into the output file.

(e) (no output, just written): 20 points.
Empirically investigate the time and memory consumption of your source code. For this, you should randomly generate pairs of DNA sequences (with alphabet ATCG) of length 50, 100, 250, 500, 1000, 1,500, 2,000, and the longest sequences that you can accomplish. Report the running time (in hh:mm:ss) and memory (in GB) usage of your code on these test data. Report the longest sequence that you did.

(f) (output: assignment1.o5): bonus question (20 extra points)
In this part, report all possible optimal alignments. In the first line write the number of optimal alignments and then put every alignment in two lines separated by an empty line.

Examples for input and output file formats:

program boucher.cpp

assignment-1.fasta:

> Sequence 1
ATCAGAGTC

> Sequence 2
TTCAGTC

assignment1.o1:
7

assignment1.o2:
 0 -2 -4 -6 -8 -10 -12 -14
-2 -1 -3 -5 -4 -6 -8 -10
-4 0 1 -1 -3 -5 -4 -6
-6 -2 -1 3 1 -1 -3 -2
-8 -4 -3 1 5 3 1 -1
-10 -6 -5 -1 3 7 5 3
-12 -8 -7 -3 1 5 6 4
-14 -10 -9 -5 -1 3 4 5

-16 -12 -8 -7 -3 1 5 3
-18 -14 -10 -6 -5 -1 3 7

assignment1.o3:
ATCAGAGTC
TTC--AGTC

assignment1.o4:
YES

assignment1.o5:
3
ATCAGAGTC
TTC--AGTC

ATCAGAGTC
TTCA--GTC

ATCAGAGTC
TTCAG--TC

ASSIGNMENT 2: MOTIF FINDING

Important notes:

- Your programs should be written in any programming language. The extension of your source code and executables will depend on your choice of programming language.
- You do not have to use linux/unix for this assignment but I recommend it
- You can choose to this assignment in partners or alone.
- The name of your program should be [your-lastname1-lastname2].[extension], e.g. boucher-ruiz.cpp. If you would like to attach a readme text file, it should be named [your-lastname1-lastname2]-readme.txt
- Put everything in a .zip and submit it via Canvas.
- As before, please give me any comments on how you found the difficulty of the assignment. It helps me prepare the next assignment.
- The DNA sequences will be given to you in a fasta file that will be specified on the command line. The output must be written in a separate output file (this simplifies the grading).

Problem Statement: In lecture, we discussed motif finding. The scalability of the algorithms we learned about are highly variable. In this assignment you will investigate this fact.

Question 1. (10 points) Write a program to generate t DNA strings, each of length n and insert a (k,d)-motif into each at a random location with replacement. Your program should take as input t, n, k, and d. It should output the t DNA strings into a fasta file. Call the program: a2_1_yourlastname.[file extension].

For example, running the program with t = 3, n = 10, k = 3, d = 1 may generate the following (motif string = TTT):

>seq1

AGATTTATAT

>seq2

CCATGTATAC

>seq3

GGGCTCTACC

2. (20 points) In Lecture 12, we learned about the "Brute Force Motif Search" (slides numbers 26 to 34). This algorithm finds and returns the set of motif strings (one for each input string) by considering all possible sets of k-mers in the set of input strings. Implement this algorithm.

Your program should take as input the input strings as a fasta file, k, and d. It should output the motif strings; if there are more than one set of motif strings, you should output one of them at random. Call the program: a2_2_yourlastname.[file extension].

For example, suppose you run the program with the following input file:

>seq1

AGATTTATAT

>seq2

CCATGTATAC

>seq3

```
GGGCTCTACC
```

and k = 3 and d = 1.

The output should be:

```
TAT

TAT

TCT
```

Question 3. (36 points) Implement the Gibbs Sampling algorithm that we learned about in Lecture 13 (slide number 82 contains the pseudocode). Call the program: a2_3_yourlastname.[file extension].

As in Question 2, your program should take as input the input strings as a fasta file, k, and d. It should output the motif strings and the score; if there are more than one set of motif strings, you should output one of them at random.

For example, suppose you run the program with the following input file:

```
>seq1

AGATTTATAT

>seq2

CCATGTATAC

>seq3

GGGCTCTACC
```

and k = 3 and d = 1.

The output should be:

```
TAT

TAT
```

Question 4. (34 points)

(a) Run all your algorithms for k = 15, d = 4, and 10 sequences of length 25, 100, and 250. Graph the time your algorithms take. Predict what will happen with longer sequences. i.e., when would each of the algorithms take longer than a week of CPU time? Put your solutions in a well-marked text, word, or pdf file.

(b) Run all your algorithms for k = 15, d = 4, and 5, 25, 50 sequences of length 25. Graph the time your algorithms take. Predict what will happen with longer sequences. i.e., when would each of the algorithms take longer than a week of CPU time? Put your solutions in a well-marked text, word, or pdf file.

ASSIGNMENT 3: K-UNIVERSAL STRING PROBLEM

Important notes:

- Your programs should be written in any programming language.
- You do not have to use linux/unix for this assignment but you are welcome to.
- Again, do this assignment individually or in pairs
- The name of your program should be [your-lastname1-lastname2].{c,cpp,java}, e.g. boucher-ruiz.cpp. If you would like to attach a readme text file, it should be named [your-lastname1-lastname2]-readme.txt.
- Put everything in a .zip and submit it via Canvas.

The purpose of this assignment is to help you learn and practice Eulerian assembly techniques that can be applied to general mathematical problems.

In class we learned about the k-universal string problem. Recall that a k-universal string is the shortest string that contains all k-length binary strings exactly once. For example, 00011101 is a 3-universal string, as it contains each of the eight binary 3-mers (000, 001, 011, 111, 110, 101, 010, 100) exactly once. Given this definition, consider the following computer science problem.

k-Universal Circular String Problem Find a k-universal circular binary string.
Input: An integer k.
Output: A k-universal circular binary string.

(a) (60 points) Write a program that solves the k-universal string problem. i.e., it will accept as input a value of k and output a k-universal string.

(b) (20 points). Give the k-universal strings for k = 4 and k = 5.   Put your output into a .docx, .txt, or .pdf file.

(c) (20 points). Are there multiple k-universal stings? Put your output into a .docx, .txt, or .pdf file.

Extra credit: (d) (25 points) Give a mathematical proof of your statement in (c). Put your output into a .docx, .txt, or .pdf file.