CMPSC 300
Bioinformatics
Spring 2021

Teach Your Colleagues About a *Bioinformatically* Incredible Research Area and Associated Tool

Figure 1: Bioinformatics comprises a lot of different areas to explore and to study. In many of these areas, there are plenty of *low-hanging fruit*, in research in which discoveries can be made, with very little effort.

Objectives

In this final project, you will be choosing a field of research and an associated tool from an area of research in Bioinformatics. You will research some of the supporting literature about your topic, investigate a prominent tool from this area and then introduce this topic and tool to your class, with some discussion. In addition, you will also write-up a report in Markdown for submission in which you discuss the area of research and the application of the tool.

Introduction

Throughout the semester, we have studied some of the many topics from Bioinformatics that are listed in Figure 1, and have also used various computational tools such as alignment and BLAST to explore a number of biological phenomena. Unfortunately, we were unable to study all topics in the field but this does not mean that your learning in the field ends with the short-list of class topics.

On the contrary, the final project invites you to explore, in greater detail, a real-world application of bioinformatics. Here, you will complete research over a topic from bioinformatics project that is interesting to you. In addition, you will pick a computational tool (online or offline) from this area that is commonly used in the study of your selected research area. The use of this tool will be demonstrated to the class, along with your discussion of the research area in which the tool is used.

HANDED OUT: 19th APRIL 2021

GitHub Starter Link for Groups

STOP! STOP!
Not everyone will be clicking this link at this time!
Only the team leader will be clicking the link to create the repository!!
https://classroom.github.com/g/K3w8yVX1

Creating your repository

If you choose, you may work in groups of up to three people for this assignment. If you would rather not work in a group, then you may work alone in a group all of your own.

Individually: If you are working alone, you will still need to establish a group name due to the setup in GitHub Classrooms.

As a group: If you are working in a group, first, decide who will be in your group before you click on the link. In group assignments, **only one person will be creating the team while the other team members will join that team.** The selected person of the team should click on the link of the assignment sheet to establish a group name that is unique and descriptive. Use the "Create a new team" option.

Now the other members of the team can click on the assignment link and select their team from the list under "Join an Existing Team." When other team members join their group in GitHub Classroom, a team is created in our GitHub organization. Every team member will be able to push and pull to their team's repository.

As you work in groups, please work collaboratively and leave no-one out of the loop of your thinking: Working in groups is one of the best ways to add focus to your project and to develop ideas (by brainstorming.)

Git commands

To push your changes, you can use the following commands to add a single file, you must be in the directory where the file is located (or add the path to the file in the command):

- `git commit <nameOfFile> -m ''Your notes about commit here''`

- `git push`

Alternatively, you can use the following commands to add multiple files from your repository:

- `git add -A`

- `git commit -m ''Your notes about commit here''`

- `git push`

Reading Assignment

General Knowledge: To help you work, please read the covered chapters from the course book, consult the week's slides and your class notes.

- **Programming**: If you are working with Python, then you can also find useful information in the Python community by performing online research for programming help.

- **Primary Sources**: For primary-source articles, please use an online search engine to locate articles which correspond to the search-keys of your work. Be sure to keep a record of the references of your work as you will need them in your written work.

- **Markdown**: Your written deliverables (are to be written in Markdown. Please take some time to gain experience with using Markdown to complete your writing assessments.

  - See *Mastering Markdown*
    `https://guides.github.com/features/mastering-markdown/` for more details about Markdown. Another good reference may be found at:
    `https://markdown-it.github.io/`.

- Be sure to read the `README.md` file in the GitHub Classroom repository for instructions on how to complete your first assignment.

- If you require other types of help to complete this work, please check with the Department of Computer Science YouTube channel which offers help in Markdown and other areas. The link is:
  `https://www.youtube.com/playlist?list=PLsYZRXov75ZHSwWiCkO-jd1RcTuu_-zmD`.

Assignment Specifications

For the project assignment you may choose a biological research question that is of the most interest to you and can be investigated through computational technique(s). Pick a research area for which a tool exists to offer some automated research. **From the literature, find about three (3) or more current research papers from which you will learn about the research area.** In research papers, tools are often mentioned as being used to procure results. In your work, choose one of these tools to demonstrate in your project. If you cannot actually run the tool, then your coverage of the tool could be spent to inform your readers about the tools, as applied to the research area.

Research: In your discussion of the research, please answer the following types of questions.

- What is the overall goal of the research?

- What disorders or problems could be resolved by a solution?

- What kind of data is necessary in this field of study?

- What did the articles that you choose say about this research area? Choose three articles to mention and provide references.

HANDED OUT: 19th APRIL 2021

Tool: In your report, you are to choose a tool to introduce. The types of questions about this tool to answer are the following.

- What is the tool?

- What is the reference (i.e., articles or website) of the tool?

- What is is used for in your research area?

- What kind of data does the tool require?

- What do the results from the tool inform us about the research area?

Ethical Component: (About 300 words, to be added to its own section of your final report document): You are to find two extra peer-reviewed research articles which could be related to your topic. Introduce these articles and use their content to help you answer the following types of questions about your research topic and tool.

- Who are the users of the technologies or methods?

- How could the tool's technology be misused?

- How could the data that the tool requires be misused?

- How could the the technologies or methods can cause harm?

- What solutions do you think should be in place to avoid or fix the harm caused by the technologies or methods?

Your project must be extensive enough to qualify as a project (think of work for at least three or four one-week lab assignments). Please focus on feasibility and scope of your project so that it can be completed by the due date. Remember, you must adhere to the Honor code! Please be original in your work!

Timeline: Deliverables

1. **Proposal** (about 300 words) Deadline: 26th April by 2:50pm; Choose an area in Bioinformatics in which you have found three peer-reviewed articles from the literature. Discuss this area in terms of the interests of the articles, described disorders or problems, required data for research, and the overall aims of the research. In addition, from this area, find a prominent automated tool (online or offline) to discuss as having an important role to play in the research. If there are other important details involved with your project, please mention them here for the instructor to consider.

   Please note: Your proposal should include at least three references from academic peer-reviewed articles to provide foundation and motivate the importance of the topic. Please also include a reference for the tool that you have selected to include in this project. *Please do not include unsubstantiated resources to your list of citations such as websites, blogs, YouTube videos, popular news articles and similar.*

2. **Progress report** (about 500 words) Deadline: Tuesday, 3rd May by 2:50pm; Describe everything you have done so far in your progress report. By this point, you should have conducted necessary research on the background of the problem, decided on the approach you will use in your research, and what significant progress towards your goals you have made and intend to make. You can also mention any challenges that you have met.

3. **Class presentation** (beginning on) Monday, 10th May during the lab session; In the presentation, you should describe the motivation, problem definition, challenges, approaches, and results and analysis from your work. Use diagrams and a few bullet points rather than long sentences and equations. The goal of the presentation is to convey the important high-level ideas and give intuition rather than be a formal specification of everything you did.

   Your presentation should be around five (5) minutes in total during which you present your slides to cover your progress. If you are working in groups, please have each member contribute to the presentation talk. At the end of the presentation give a demonstration of the tool, or showcase how the tool would be used in the area of your topic.

   Note: since your report is due *after* your talk, your talk does not need to conclude the project, rather it should provide a very informed series of events to describe your work.

4. **Final report** (About 2000 words. Note, you will also be including graphics, screen shots and etc.) Deadline: 20th May by 9:00am; Incorporate any feedback from the progress report and the presentation session. Your final report should be clear, concise and, most importantly, well written, this includes no typos or grammatical errors. Your report should be written in a professional manner and should include the following main sections.

   - The motivation for your project. Why is the topic you decided to address important and useful?

   - Background for the selected research topic. What have others in this area already? Describe the nature of the research. Use and reference your articles. See the above questions to help guide your discussion.

   - What is the tool you selected? What data is necessary, What does this tool do? How can this tool offer insight? Can you offer plots, screenshots or other visual aids to help understand how the tool functions? Please see the above questions about the tool to guide your discussion.

   - Conclusion. Give a short overview of your project and the tool that is used in the area. What should your reader know about the effectiveness of this research with the tool?

   - The ETHICAL component will be an additional 300 words.

Required Deliverables

- Markdown file: `writing/proposal/proposal.md`

- Markdown file: `writing/progressReport.md/progressReport.md`

- Markdown file: `writing/finalReport/finalReport.md`

HANDED OUT: 19th APRIL 2021

Grading

checkmark **Proposal**

checkmark **Progress report**

20 points: **Presentation and demonstration**

50 points: **Final report and Tool demonstration**

30 points: **Ethical component**

Submission Guidelines

For each deliverable, you are to submit a Markdown file of your report (or presentation slides). For your final report you are to submit any necessary and supplementary material. This includes programs, data sets, a *README.txt* file documenting what everything is (i.e., a justification of the existence of the files that you have left for the instructor in your repository). Finally, for your code, you will need to write up documentation to instruct how the code is to be used and what its expected inputs and outputs should be. Please write up these details so that they will be useful to the instructor.

In adherence to the Honor Code, students should complete this assignment while exclusively collaborating with the other member of their team. While it is appropriate for students in this class who are not in the same team to have high-level conversations about the assignment, it is necessary to distinguish carefully between the team that discusses the principles underlying a problem with another team and the team that produces an assignment that is identical to, or merely a variation on, the work of another team. Deliverables from one team that are nearly identical to the work of another team will be taken as evidence of violating Allegheny College's Honor Code. Do not be tempted to look online for possible problems and solutions, that institutes a violation to the Honor code! Please be original!