AUG
Faculty of Computer and Information Sciences
Ain Shams University

Session 1
Orientation Session : Hello Bioinformatics

This is the first session of the first level of our program "Problem-Solving for Bioinformatics". We are aiming to help you learn more about the field of bioinformatics through a series of problem solving challenges, so you get to learn about the amazing field of Bioinformatics while strengthening your problem solving skill set.

Layout & Rules

- The program currently consists of 2 levels with plans undergoing to add more.
- Each level is scheduled to take place over 3 months, so you get to learn at a reasonable pace.
- Each level consists of 8 main sessions, a session every week for two months with the third month left for project development & support sessions on topics you decide.
- You MUST attend ALL of our online sessions on our Microsoft Teams official channel.
- Every session will be in the span of 60 - 90 mins with extra support time for those who ask/need.
- There will be an online form to submit any questions regarding the sessions or the regulations, suggestions & complaints.
- Sessions might be recorded.
- You will be evaluated over session attendance, participation, task/project delivery & last but not least integrity.

What is Bioinformatics?

In simple words it's the application of information & computational technologies in relation to molecular biology. The goal is to catalyze solving complex problems in the biological field with the arsenal that computer & information sciences can afford.

So we can say that Bioinformatics is an interdisciplinary field between biology & computer science.

In this program we will study some concepts of Bioinformatics through problem solving. The technology/language we will use is python.

Why python?

To answer this question we will have to take a quick dive into some tech concepts. We will mainly try to differentiate between Development & Scripting then Procedural & Declarative programming.

A) Development & Scripting

Development	Scripting
Aims to create full systems that consist of instructions then fed some input to generate a specific output for every case. It's associated with developing large software tools that help to solve problems or provide services to its users. Defined as the main aim of industry, so efficiency, ease of use, scalability, responsiveness, reliability, positive experience are all considered in development.	Task specific code that aims to solve a problem in a defined context just to obtain the resulting output. Aspects that are cared about are efficiency, accuracy & reproducibility. In general, most of the time scripts are subsets of larger organized systems that manage their workflow. Stand-alone scripts are popular in scientific research.
Uses compiled languages.	Relies on interpreted languages
C, C#, C++, Java, Scala.	Python, R, Javascript, perl

So as we can observe, since we are going to learn bioinformatics by designing simple solutions to fundamental bioinformatics problems, we are going to rely on scripting. Each script shall solve a certain problem.

B) Procedural & Declarative Programming

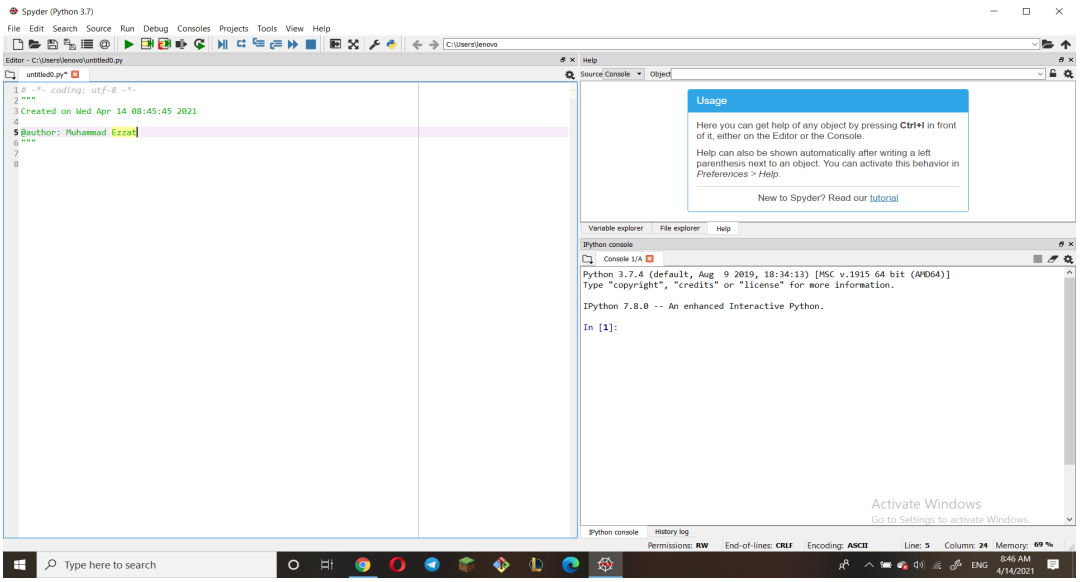
Both of them are useful programming paradigms but they aren't the only ones. You can find a comprehensive list of programming paradigms from here. However, we chose to discuss those two as you probably haven't been exposed to any but those two.

Procedural Programming	Declarative Programming
Is telling the computer exactly how to do a task/job in steps/procedures. Procedural programming defines the entire workflow of a machine.	Is telling the computer just what to do. You don't need to specify the workflow.
C, C++, Python	SQL, R, Python

What we can understand is that python can combine both paradigms to deliver you the best of the two worlds. While we design our solution we will mostly care about defining the workflow of the scripts core, but for some other surrounding tasks it will be easier to rely on the declarative end of programming.

Tools

There are many tools for writing python, however we recommend using spyder as It's a robust environment & empowers scripting.



However, feel free to use the tool you are comfortable with. Some of the alternatives are: pyCharm, Jupyter, Eclipse, Atom...etc

Continuous Project :

You will be asked to develop a project along the course. Don't worry it will be all related to the studied content & we will keep it simple. Plus mentoring is always available.

For level 1 : Sequence Wizard !

Input : A Fasta file containing one or more DNA sequence(s).

Output :

- Nucleotide Count & Plot
- Complementary sequence
- Corresponding RNA sequence
- Corresponding Protein sequence
- GC content ratio
- GC skew plot
- K-mers of the given sequence (for given K)
- Finding a certain K-mer
- K-mer Frequency
- Popular K-mers of X mismatches (for given K,X)

Don't feel intimidated you'll learn how to solve each subproblem during the program. We will ask you to deliver the project at the end of level 1.

Introduction to python

Please refer to the associated code files (.py).