

Homework 1

1. In cpts317, we learned that a problem is a language. Such a statement builds a link from 317 to 350. For instance, consider the following problem:

Given: a graph G ,

Question: is G connected?

This problem corresponds to the language that puts all the string encodings of the positive instances of the problem into: $\{\langle G \rangle : G \text{ is a connected graph}\}$, where $\langle G \rangle$ is the string encoding of G . Please indicate the languages corresponding to the following four problems:

Given: a number n and two primes p and q ,

Question: is it the case that $n = p \cdot q$?

Given: a number n ,

Question: is it the case that $n = p \cdot q$ for some primes p and q ?

Given: an NFA A and a word w ,

Question: Does A accept w ?

Given: an NFA A ,

Question: Is there a word w such that A accepts w ?

2. Prove why the following statements are true:

(1). Function $2n^3 - 18n$ is $O(n^3)$ and also it is $O(n^4)$ but is it not $O(n^2 \log n)$.

(2). Function $3n^2 2^{2n}$ is $2^{O(n)}$.

3. (You MUST do this problem since your TA, very likely, may choose to grade this problem only. It is also good exercise to think through real world problems while learning algorithms.) In designing an algorithm for a real world complex problem, probably the most difficult part is to figure out (a) input to your algorithm including its data structure so that you may bring the input in your memory and feed it to your algorithm; (b) what the problem really is (without understanding the problem, you can't even design an algorithm). In class, I went through a small example. Now, you are going to work on one more example and write an essay about it. First, stare at a protein 3D structure (which is one big molecule) shown in the following link for 5 minutes

<https://www.rcsb.org/3d-view/4hhb/1>

You may move your mouse over the picture to see the forming atoms and bonds. Now, write to answer the following questions — herein, the final goal for you is to design an algorithm M to compute a similarity metric (which is a number and *you* define it!) between two protein molecules.

A. Read from the Internet and write a not-so-short paragraph on why such an algorithm M (or related ideas) could help scientists to develop new medical drugs to fight with tough diseases in the future.

B. Since the input to your similarity algorithm M is of two protein molecules as shown in the link above, as computer scientists, we first need consider how to represent a protein molecule in computer memory; i.e., choosing a data structure for it. Of course, you have many ways to do it. For instance, we can represent it as a text string, as a graph, or even as a jpeg picture. You describe a way on how to represent a protein molecule.

C. You sketch two definitions of the similarity metric between two protein molecules and also describe the Pros and Cons of the two definitions. You also need sketch two algorithms (and intuitively analyze their efficiencies) M (over the data structure that you described in B above) for the two definitions, respectively.

This is a world class problem; so even though your similarity metric definitions and algorithms are not so sophisticated, you can still obtain a high score. The purpose of this problem is to force yourself to go through the problem solving process by designing algorithms and may attract yourself into bioinformatics (If you are interested in these topics, you may want to take future courses in

Bioinformatics, Data Science, Machine Learning, Neural Networks, etc. Remember that you don't have to be a biologist in order to do bioinformatics and in order to find a high-pay computer science job at the biggest pharmaceutical companies like Merck, Pfizer, etc., in the world.).

Feel free to read off the Internet and cite the source when you borrow others' ideas.