BT 3051 — Data Structures and Algorithms for Biology
Jul–Nov 2019
Course Plan

Karthik Raman
Department of Biotechnology
Indian Institute of Technology Madras

Lectures

Lectures:	**E Slot**, Tuesdays (11:00), Wednesdays (10:00)
Labs:	**T Slot**, Fridays (14:00)*
Tutorial:	**E Slot**, Fridays (16:50)
Venue:	BT 108
E-mail:	kraman @ iitm·ac·in
Office:	BT 221 (Block II)
Office hours:	By appointment
Teaching Assistants:	Lavanya Raajaraam (bt17d401@smail / BT-II 201)
	Shreya Swaminathan (be15b027@smail / BT-II 201)
	Shreyansh Umale (be15b028@smail)
	Ramya Vijayram (be15b032@smail)
Piazza:	http://piazza.com/iitm.ac.in/fall2019/bt3051

*Bring your laptop! We will try to stick to the Fri lab schedule, but it may not always happen; so check your emails before you come to class!

Objectives

- Introduce you to basic algorithms (methods for problem solving)
- Introduce elementary data structures (methods for storing information)
- Study algorithms and data structures of particular importance in biology
- Improve your programming skills (learn good programming practices)

At the end of the course, it is expected that you will

- have a good understanding of basic algorithms/data structures
- understand and apply general computational techniques such as dynamic programming/randomisation/…
- be able to understand and use standard libraries to solve biological problems
- be able to develop algorithms and data structures to solve biological problems
- … and write and test correct and readable programs!
- inculcate professionalism

Piazza

- This term, we will be using Piazza for class discussion
- The system is highly catered to getting you help fast and efficiently from classmates, the TAs, and myself

    – Piazza is of course the anti-thesis of RG!

- Rather than emailing questions to me/TAs, I encourage you to post your questions on Piazza
- In fact, every question you ask me after class, I would encourage you to post it on Piazza for the benefit of your classmates
- Class page: https://piazza.com/iitm.ac.in/fall2019/bt3051/home
- If you have any problems or feedback for the developers, email team@piazza.com

Course contents

1. Introduction + Basics of Python/Programming
2. Introduction to Algorithms and Data Structures
3. Sorting algorithms, Dynamic Programming
4. String algorithms
5. Graph algorithms
6. Random numbers, sampling
7. Direct search algorithms / Evolutionary algorithms

Pre-requisites

There are no formal pre-requisites for the course, other than a keen interest in programming, problem-solving (and biology)!

Coursework and grading

- Attendance to the lectures is required; *late-comers will be marked absent*
- You must bring your laptops to the lab; you will not be allowed to attend the lab without a laptop
- Weightage (minor changes may happen):

    – Quizzes I/II/Pop: 30%

    – Homework: 30%

    – End-semester exam: 40%

- Some homework may be assigned in pairs
- Class participation (incl. on Piazza online) will also carry weightage ($\approx$bonus)
- Grading will be absolute

For every hour spent in class, I expect you to spend at the least <u>3 hours outside of class</u> on reading relevant and additional material/programming to understand the concepts. I also expect you to inculcate professionalism and maintain the highest level of integrity.

Important Dates

Date	Day	Time	Event
5-Sep-19	Thu	08:00	Quiz I
13-Sep-19	Fri	11:00	Mid-term feedback
10-Oct-19	Thu	08:00	Quiz II
30-Oct-19	Wed	11:00	Final feedback/TCF
21-Nov-19	Thu	09:00	End-semester Exam

Exams

The exams will be as per Institute timetable: Quiz I will be on 5th September 2019, while Quiz II will be on 10th October. The final exam will be on 21st November.

Homework

There will be 5 assignments during the course. They will mostly involve programming in Python. Credit will be given for correctness, and importantly, also for coding style, comments, test cases, readability. Please have a look at this page (from Cosma Shalizi's blog), on minimal advice to under-graduates on programming: http://cscs.umich.edu/~crshalizi/weblog/593.html.

All homework will be due at 5 pm on the day mentioned in the problem handout, via a Dropbox upload at http://tinyurl.com/bt3051-submit. Late submissions will incur penalties as follows:

- 1 second – 24 h: 20%
- 24–48 h: 40%
- > 48h: 60%

Academic integrity

While you can take help from your colleagues on homework, *copying* is not permitted. Copying from anywhere, including the Web is not allowed. Offenders will be penalised letter grades. For example, you can ask your friends for hints if you are stuck, but at no point should you take a look at their code itself. I would much rather that you come to me if you are stuck. Note that I take copying extremely seriously, and will not show any mercy (ask your seniors!). Many students have failed this course in the past owing to copying of assignments.

The Institute guidelines (which I consider lenient) may be found here: http://academic.iitm.ac.in/sites/default/files/Graded_punishments.pdf. You must not carry your mobile phone/e-book reader to the exam desk, even if they have been switched off.

Feedback

Any suggestions for improvement are welcome at any time. A mid-term feedback will happen on 14th September. The final feedback will happen on 30th October.

Reading

There are many nice books on algorithms, but none ideally suited to our course, to discuss biologically relevant algorithms. Some of the books below are excellent:

- Gries P, Campbell J, and Montojo J (2013) *Practical Programming: An Introduction to Computer Science Using Python 3 (Pragmatic Programmers).* Pragmatic Bookshelf, third edition/e. ISBN 9789351104698

- Goodrich MT, Tamassia R, and Goldwasser MH (2013) *Data Structures and Algorithms in Python.* Wiley, 1/e. ISBN 1118290275

- Compeau P and Pevzner P (2014) *Bioinformatics Algorithms: An Active Learning Approach.* Active Learning Publishers, 1st/e. ISBN 0990374602

- Guttag JV (2013) *Introduction to Computation and Programming Using Python.* The MIT Press, revised and expanded edition/e. ISBN 0262525003

- Skiena SS (2010) *The Algorithm Design Manual.* Springer, softcover reprint of hardcover 2nd ed. 2008/e. ISBN 1849967202

- Cormen TH (2010) *Introduction to algorithms.* PHI Learning. ISBN 9788120340077

- Sedgewick R, Wayne K, and Dondero R (2015) *Introduction to Programming in Python: An Interdisciplinary Approach.* Addison-Wesley Professional, 1/e. ISBN 0134076435

Tentative schedule

Week	Date	Day	Lecture #	Lab #	Topic
1	30-Jul-19	Tue	0		Administrivia
1	31-Jul-19	Wed	1		Diagnostic Quiz 0
1	2-Aug-19	Fri	2		Overview
1	2-Aug-19	Fri		1	Lab: Intro to Python
2	6-Aug-19	Tue		2	Lab: Intro to Python
2	7-Aug-19	Wed	3		Introduction to Algorithms
2	9-Aug-19	Fri		3	Lab: Intro to Python
2	9-Aug-19	Fri		4	Lab: Timing algorithms
3	13-Aug-18	Mon	4		Introduction to Python
3	13-Aug-19	Tue	5		Analysis of Algorithms / Order of growth Classifications
3	14-Aug-19	Wed	6		Order of growth Classifications / Asymptotic Notation
3	16-Aug-19	Fri		5	Lab: Timing algorithms
3	16-Aug-19	Fri		6	Lab: Introduction to Object-oriented programming
4	20-Aug-19	Tue	7		Introduction to Data Structures/Basics of Data structures
4	21-Aug-19	Wed	8		Fundamental Data Structures
4	23-Aug-19	Fri	9		Basic Data Structures
5	27-Aug-19	Tue	10		More Data Structures
5	28-Aug-19	Wed	11		Elementary Sorting algorithms
5	30-Aug-19	Fri		7	Lab: Implement Some Basic Data Structures
5	30-Aug-19	Fri		8	Lab: Evaluate an arithmetic expression
6	3-Sep-19	Tue	12		Divide and conquer: Mergesort
6	4-Sep-19	Wed	13		Introduction to Dynamic programming
6	5-Sep-19	Thu			**Quiz I**
6	6-Sep-19	Fri		9	Lab: Data Structures
6	6-Sep-19	Fri		10	Lab: Data Structures / Sorting
7	10-Sep-19	Tue			Holiday – Muharram

Week	Date	Day	Lecture #	Lab #	Topic
7	11-Sep-19	Wed			Holiday – Onam
7	13-Sep-19	Fri		11	Lab: Implement sorting algorithms
7	13-Sep-19	Fri		12	Lab: Implement sorting algorithms; **Mid-term course feedback**
8	17-Sep-19	Tue	14		Dynamic programming
8	18-Sep-19	Wed	15		Dynamic programming wrap-up
8	20-Sep-19	Fri		13	Lab: Dynamic Programming
8	20-Sep-19	Fri		14	Lab: Dynamic Programming
9	24-Sep-19	Tue	16		Introduction to String matching algorithms
9	25-Sep-19	Wed	17		String matching algorithms
9	27-Sep-19	Fri		15	Lab: Regular Expressions
9	27-Sep-19	Fri		16	Lab: String matching
10	1-Oct-19	Tue	18		Regular Expressions
10	2-Oct-19	Wed			Holiday – Gandhi Jayanti
10	4-Oct-19	Fri		17	Lab: Parsing JSON and XML/Regular Expressions
10	4-Oct-19	Fri		18	Lab: NetworkX
11	8-Oct-19	Tue			Holiday – Vijayadashami
11	9-Oct-19	Wed	19		Tries
11	10-Oct-19	Thu			**Quiz II**
11	11-Oct-19	Fri	20		Suffix Trees
11	11-Oct-19	Fri	21		Introduction to graphs
12	15-Oct-19	Tue	22		Graph algorithms
12	16-Oct-19	Wed	23		Graph algorithms
12	19-Oct-17	Thu	24		Graph algorithms
12	18-Oct-19	Fri		19	Lab: NetworkX
12	18-Oct-19	Fri		20	Lab: Graph algorithms
13	22-Oct-19	Tue	25		Random number generation
13	23-Oct-19	Wed	26		Random sampling
13	25-Oct-19	Fri		21	Lab: Graph algorithms
13	25-Oct-19	Fri		22	Lab: Random number generation
14	29-Oct-19	Tue	27		Random sampling
14	30-Oct-19	Wed	28		Random sampling / Bootstrapping; **TCF / Final course feedback**
14	1-Nov-19	Fri	29		Introduction to Direct search algorithms
14	1-Nov-19	Fri	30		Evolutionary algorithms
15	5-Nov-19	Tue	31		Introduction to DNA Computing
15	6-Nov-19	Wed	32		Buffer class
15	8-Nov-19	Fri		23	Lab: Random number generation
15	8-Nov-19	Fri		24	Lab: Random sampling / Bootstrapping
16	12-Nov-19	Tue			Holiday – Guru Nanak Jayanti
16	13-Nov-19	Wed	33		*Informal class: Q & A*
18	26-Nov-18	Mon			**End Semester Examination**

Last updated: July 26, 2019