BT 3051 — Data Structures and Algorithms for Biology
Jul–Nov 2014
Course Plan

Karthik Raman
Department of Biotechnology
Indian Institute of Technology Madras

Lectures

Lectures:	**D Slot**, Mondays (11:00), Tuesdays (10:00)
Labs:	**D Slot**, Wednesdays (09:00), Thursdays (13:00)*
Venue:	BT 108
E-mail:	kraman @ iitm·ac·in
Office:	BT 104
Office hours:	By appointment
Teaching Assistants:	Aarthi R (bt13d031@smail)
	Aditya Pratapa (bt13s005@smail)
Piazza:	http://piazza.com/iitm.ac.in/fall2014/bt3051

*Bring your laptops!

Objectives

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

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 develop algorithms and data structures to solve biological problems
- … and write and test correct and readable programs!
- inculcate professionalism

Piazza (experiment!)

- 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/fall2014/bt3051/home
- If you have any problems or feedback for the developers, email team@piazza.com

Course contents

1. Introduction + Basics of Python/Programming (5 lectures + 4 labs)
2. Introduction to Algorithms and Data Structures (5 lectures + 4 labs)
3. Sorting algorithms, Dynamic Programming (4 lectures + 2 labs)
4. String algorithms (3 lectures + 2 labs)
5. Graph algorithms (3 lectures + 2 labs)
6. Hidden Markov Models, Markov Chain Monte Carlo (4 lectures + 2 labs)
7. Random numbers, sampling, Metropolis algorithm (4 lectures + 3 labs)
8. Evolutionary algorithms (2 lectures + 1 lab)
9. Miscellaneous topics (3 lectures + 2 labs)

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
- 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 2–3 hours outside of class 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
9-Sep-14	Tue	08:00	Quiz I
16-Sep-14	Tue	10:00	Mid-term feedback
16-Oct-14	Thu	08:00	Quiz II
10-Nov-14	Mon	11:00	Final feedback/TCF
24-Nov-14	Mon	09:00	End-semester Exam

Exams

The exams will be as per Institute timetable: Quiz I will be on 9th September 2014, while Quiz II will be on 16th October. The final exam will be on 24th November.

Homework

There will be 6–8 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.

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.

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.

Tentative schedule

Week	Date	Day	Lecture #	Lab #	Topic
1	30-Jul-14	Wed	0		Administrivia
1	31-Jul-14	Thu	1		Overview
2	4-Aug-14	Mon	2		Introduction to Python
2	5-Aug-14	Tue	3		Python Basics
2	6-Aug-14	Wed		1	Lab: Familarise with Python
2	7-Aug-14	Thu		2	Lab: Simple python programs
3	11-Aug-14	Mon	4		Data Structures in Python
3	12-Aug-14	Tue	5		Programming in Python
3	13-Aug-14	Wed		3	Lab: Understanding Python Data Structures
3	14-Aug-14	Thu		4	Lab: Finding palindromic DNA
4	18-Aug-14	Mon	6		Introduction to Algorithms
4	19-Aug-14	Tue	7		Analysis of Algorithms
4	20-Aug-14	Wed		5	Lab: Timing algorithms
4	21-Aug-14	Thu		6	Lab: Programming Style/Debugging/Testing
5	25-Aug-14	Mon	8		Data Structures: Introduction
5	26-Aug-14	Tue	9		Basic Data Structures
5	27-Aug-14	Wed		7	Lab: Implement Some Basic Data Structures
5	28-Aug-14	Thu		8	Lab: Evaluate an arithmetic expression
6	1-Sep-14	Mon	10		More Data Structures
6	2-Sep-14	Tue	11		Sorting algorithms
6	3-Sep-14	Wed		9	Lab: Implement sorting algorithms
6	4-Sep-14	Thu	12		Divide and conquer: Mergesort

Week	Date	Day	Lecture #	Lab #	Topic
7	8-Sep-14	Mon	13		Introduction to Dynamic programming
7	9-Sep-14	Tue			**Quiz I** (B slot)
7	10-Sep-14	Wed	14		Dynamic programming
7	11-Sep-14	Thu		10	Lab: Dynamic Programming
8	15-Sep-14	Mon	15		String matching algorithms
8	16-Sep-14	Tue	16		String matching algorithms, Suffix Trees; **Mid-term course feedback**
8	17-Sep-14	Wed	17		Regular Expressions and Parsing
8	18-Sep-14	Thu		11	Lab: Regular Expressions and Parsing
9	22-Sep-14	Mon	18		Introduction to graphs
9	23-Sep-14	Tue	19		Graph algorithms
9	24-Sep-14	Wed		12	Lab: Parsing BLAST XML/Biopython/BeautifulSoup
9	25-Sep-14	Thu		13	Lab: Numpy/Scipy/Matplotlib
10	29-Sep-14	Mon			No class – Friday Timetable
10	30-Sep-14	Tue	20		Graph algorithms
10	1-Oct-14	Wed		14	Lab: NetworkX
10	2-Oct-14	Thu			Holiday – Gandhi Jayanti
11	6-Oct-14	Mon			Holiday – Bakrid
11	7-Oct-14	Tue	21		Hidden Markov Models
11	8-Oct-14	Wed	22		Hidden Markov Models
11	9-Oct-14	Thu		15	Lab: Hidden Markov Models
12	13-Oct-14	Mon	23		Markov Chain Monte Carlo
12	14-Oct-14	Tue	24		Markov Chain Monte Carlo
12	15-Oct-14	Wed		16	Lab: Markov Chain Monte Carlo
12	16-Oct-14	Thu			**Quiz II** (E slot)
13	20-Oct-14	Mon	25		Random number generation
13	21-Oct-14	Tue	26		Random number generation
13	22-Oct-14	Wed			Holiday – Deepavali
13	23-Oct-14	Thu		17	Lab:Random number generation
14	27-Oct-14	Mon	27		Random sampling
14	28-Oct-14	Tue	28		Metropolis algorithm
14	29-Oct-14	Wed		18	Lab: Simulated annealing
14	30-Oct-14	Thu		19	Lab: Random walks
15	3-Nov-14	Mon	29		Evolutionary algorithms
15	4-Nov-14	Tue	30		Evolutionary algorithms
15	5-Nov-14	Wed		20	Lab: Evolutionary Algorithms
15	6-Nov-14	Thu		21	Lab: Miscellaneous topics
16	10-Nov-14	Mon	31		Miscellaneous topics; **Final feedback/TCF**
16	11-Nov-14	Tue	32		Miscellaneous topics
16	12-Nov-14	Wed		22	Lab: Miscellaneous topics
16	13-Nov-14	Thu	33		DNA Computing
17	17-Nov-14	Mon			*Informal class: Q & A*
18	24-Nov-14	Mon			**End Semester Examination**

Feedback

Any suggestions for improvement are welcome at any time. A mid-term feedback will happen on 16th September, 2014. The final feedback will happen on 10th November.

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:

Books

- Gries P, Campbell J, and Montojo J (2013) *Practical Programming: An Introduction to Computer Science Using Python 3 (Pragmatic Programmers)*. Pragmatic Bookshelf, 2/e. ISBN 9789351104698
- Jones NJ and Pevzner PA (2009) *Introduction to Bioinformatics Algorithms*. ANE Books. ISBN 8180520781
- Guttag JV (2013) *Introduction to Computation and Programming Using Python*. The MIT Press, revised and expanded edition/e. ISBN 0262525003
- Cormen TH (2010) *Introduction to algorithms*. PHI Learning. ISBN 9788120340077
- Skiena SS (2010) *The Algorithm Design Manual*. Springer, softcover reprint of hardcover 2nd ed. 2008/e. ISBN 1849967202
- Heineman GT and Pollice G (2008) *Algorithms in a Nutshell*. Shroff Publishers & Distributors Pvt Ltd. ISBN 8184046081

Last updated: August 1, 2014