

CSCI 2110

Data Structures and Algorithms

Dr. Srini Sampalli

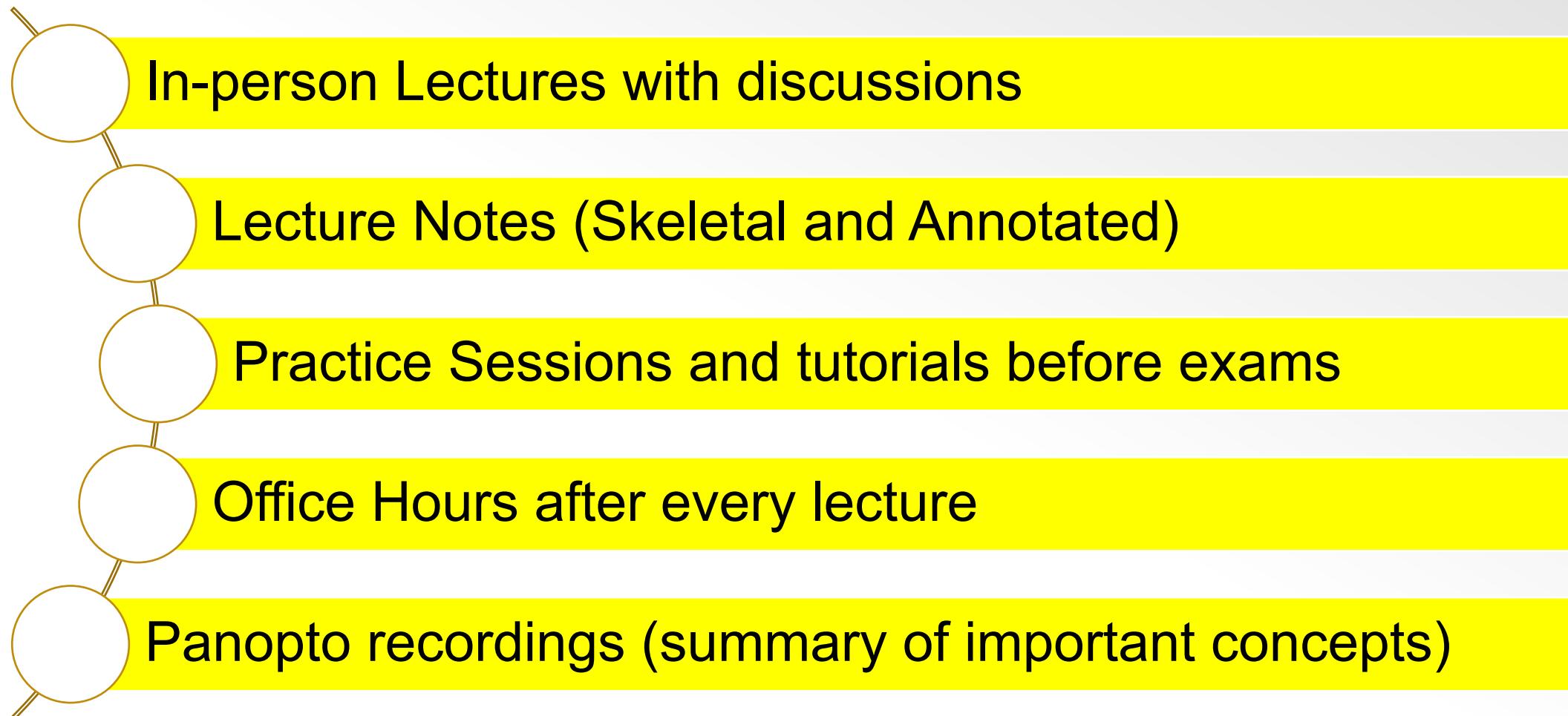
WELCOME!



**DALHOUSIE
UNIVERSITY**

*Welcome back to
in-person classes!*

Course Delivery Structure



Why it is important to attend live lecture sessions...

- You will get more detailed explanations.
- You will have the opportunity for engaging better.
- You will not be putting off understanding concepts.
- I will provide “Cyber Info Bytes” at the beginning of each lecture.
- Pop quizzes for the course will be held during the lecture times.
- **Students who attend at least 70% of lectures in person will get a 2% bonus!**

Live Streaming of Lectures via Collaborate Ultra

- Lectures will be live-streamed on Collaborate Ultra.
- Use this only occasionally, when you are unable to attend and only for exceptional circumstances.
- Please note that the quality of the delivery will depend upon the network connectivity and other issues -> cannot be guaranteed.

Exceptional situations for attending remotely

- Medical Reasons
- Family/ Personal Emergencies
- Weather situations
- Temporary travel restrictions
- Academic Accommodations

Reasons that do not qualify for attending remotely!

- “It is a 10.30 AM lecture – can’t wake up that early!”
- “I have a conflicting schedule”
- “I am in a co-op work term”
- “Personal convenience - I cannot commute to campus”

How to get the most out of the course...

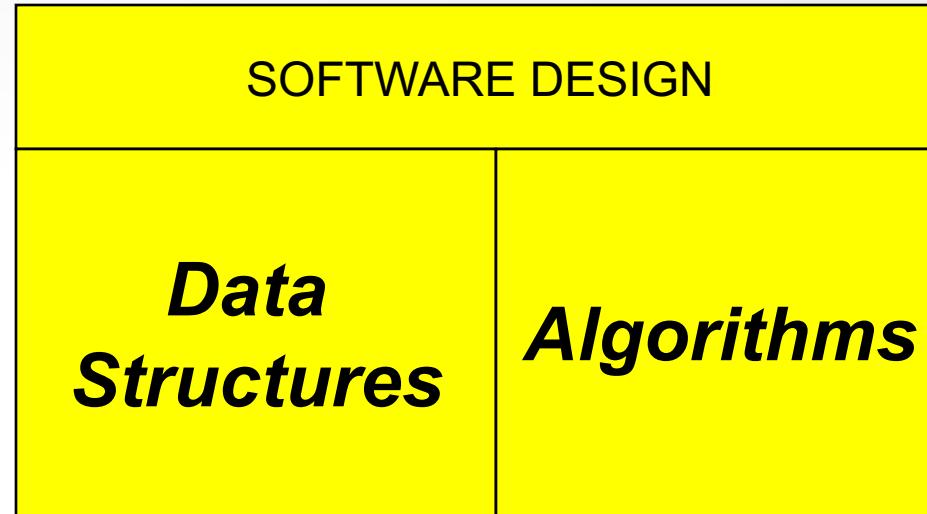
- Preview skeletal notes, attend and engage in lecture sessions, review annotated notes.
- Try to avoid multitasking during the lecture.
- Be organized – lecture notes, labs, assignments.
- Commit study time to review after the lecture.
- **Explore further!**

CSCI 2110 Data Structures and Algorithms: Course Overview

- Motivation
- Course Objective
- Topic Outline
- Grading Scheme
- Course Logistics

Motivation for learning Data Structures and Algorithms

*“If programs are the question,
then data structures and algorithms are the answer.”*



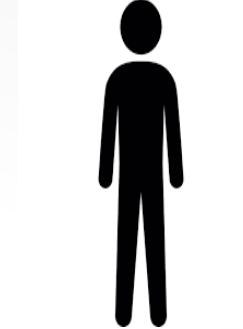
- *Need to choose the right data structure for a given problem.*
- *Need to choose the right algorithm for that data structure.*

We are living in BIG DATA times

- There are over 5.6 billion searches performed on the Internet each day, 228 million each hour, 3.8 million each minute.
- As of April 2023, there are 5.18 billion Internet users, 4.65 billion social media users and 7.6 billion mobile phone users worldwide (the world's population is 8 billion).
- 270,000 text messages are sent every second.
- Facebook has close to 3 billion active users, up from 1 million in 2004 (If Facebook was a country, it would be the third most populous).
- Google has indexed over 130 trillion web pages, up from 1 trillion in 2008.
- In the next five years, there will be 50 billion smart devices that can collect, analyze and share data.

Sources: statista.com and forbes.com

**The exploding digital universe:
1.7 megabytes of
new information
is generated
every second
for every human being on this
planet.**



*We need to understand and apply
efficient **data structures** and smart
algorithms that work on **large data!***

A Sampling of Applications

Medicine and Molecular Biology: DNA Analysis

Researchers are most interested in *finding patterns* in DNA sequences.

Examples:

Find the longest common subsequence in the two sequences

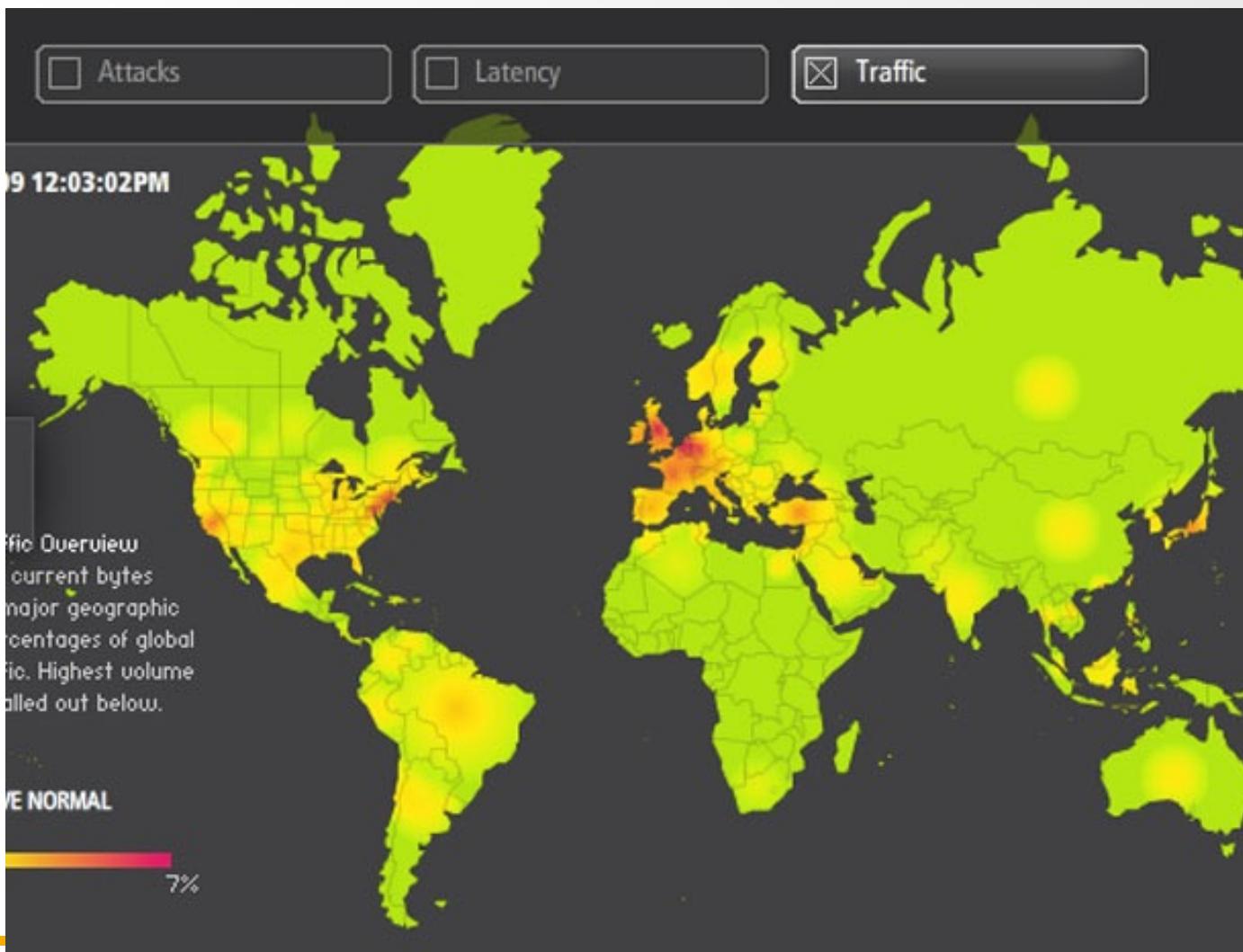
GAAGGTCGAA and CTCGGGGA

If two DNA sequences have a large subsequence in common, they often show similar function.

Detect a DNA Palindrome

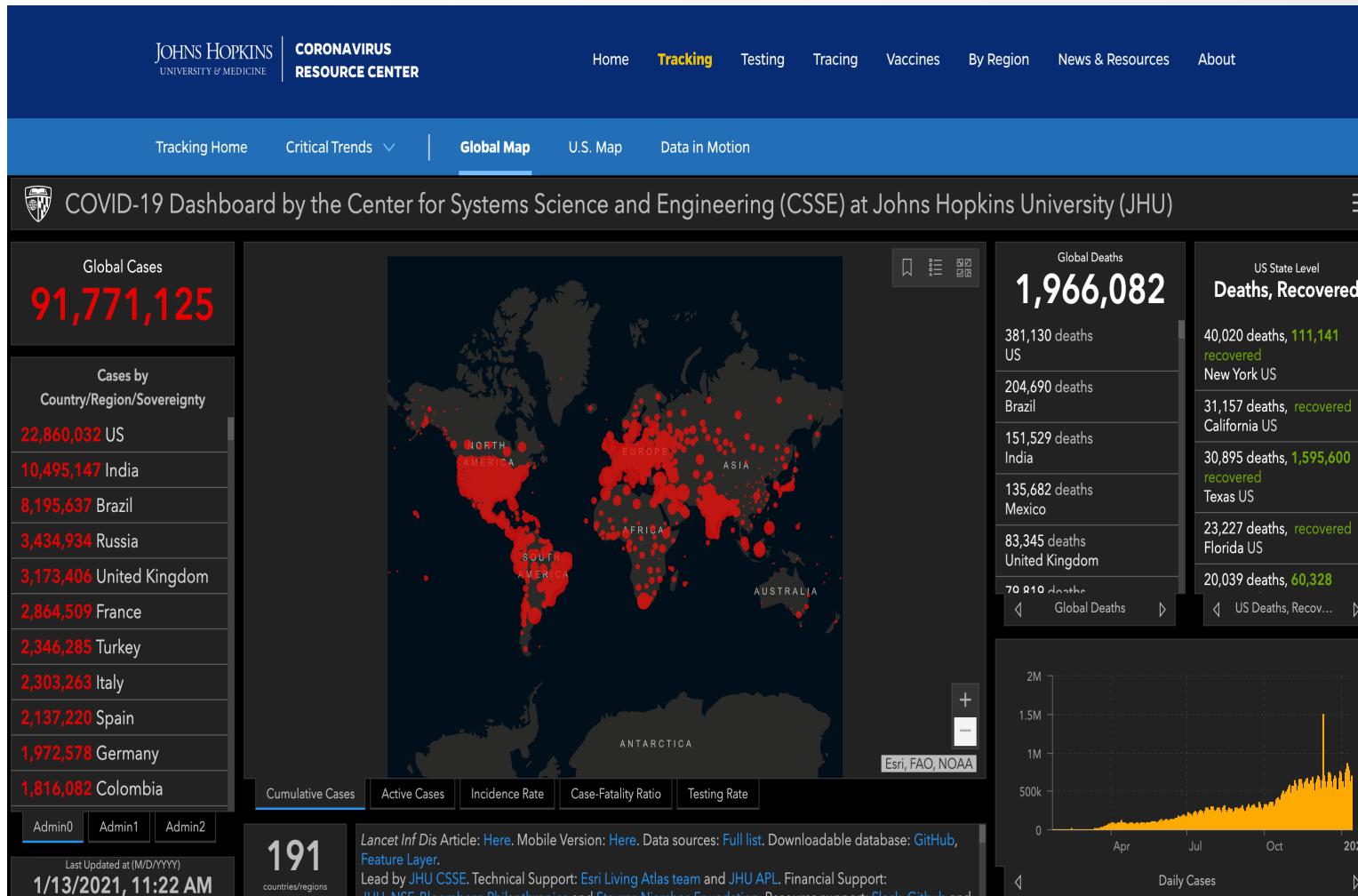
It is fascinating that something as simple as detecting a DNA palindrome might help detect what kinds of cells are cancerous or have potential to become cancerous.

Networking: Monitoring Web Traffic for Latency and Cyber Attacks



Source: Akamai Networks

Monitoring real-time Covid-19 data

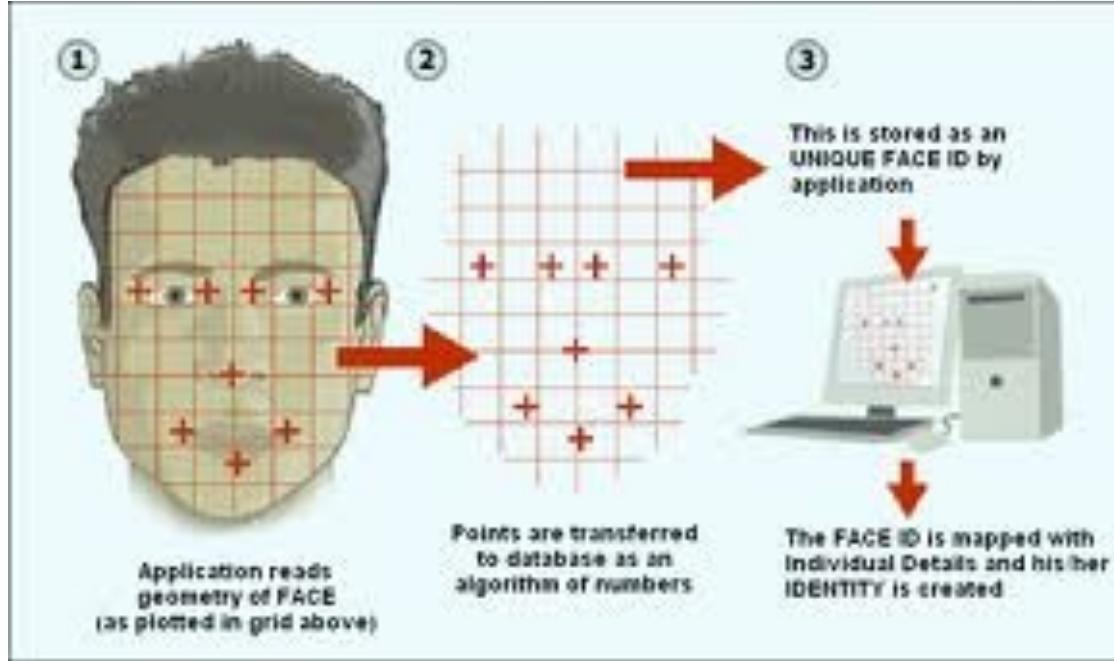


Algorithms for Data Analytics:

- 1. Which countries/regions flattened the curve the fastest?*
- 2. Which countries/regions had the highest mortality rates?*
- 3. Which countries/regions had the highest vaccine doses administered per 100 people?*

Source: <https://coronavirus.jhu.edu/map.html>

Biometrics: Face Recognition Software

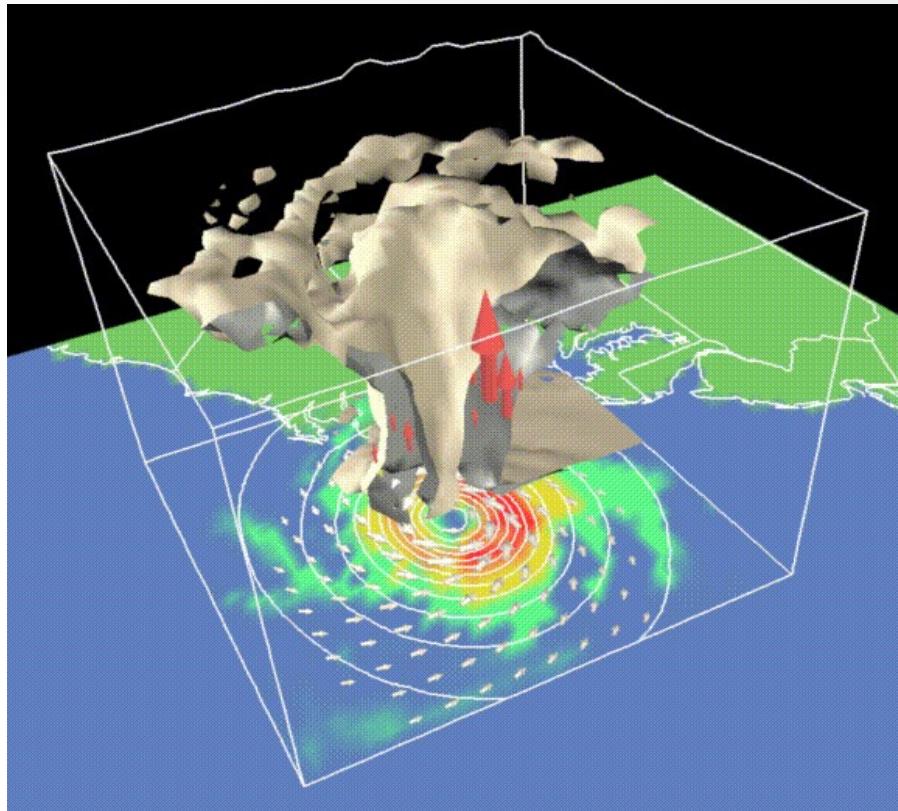


**Given a face,
match it with a
template
- relatively easy**

**Given a sea of faces, identify if a face is
one from the template – really hard!**



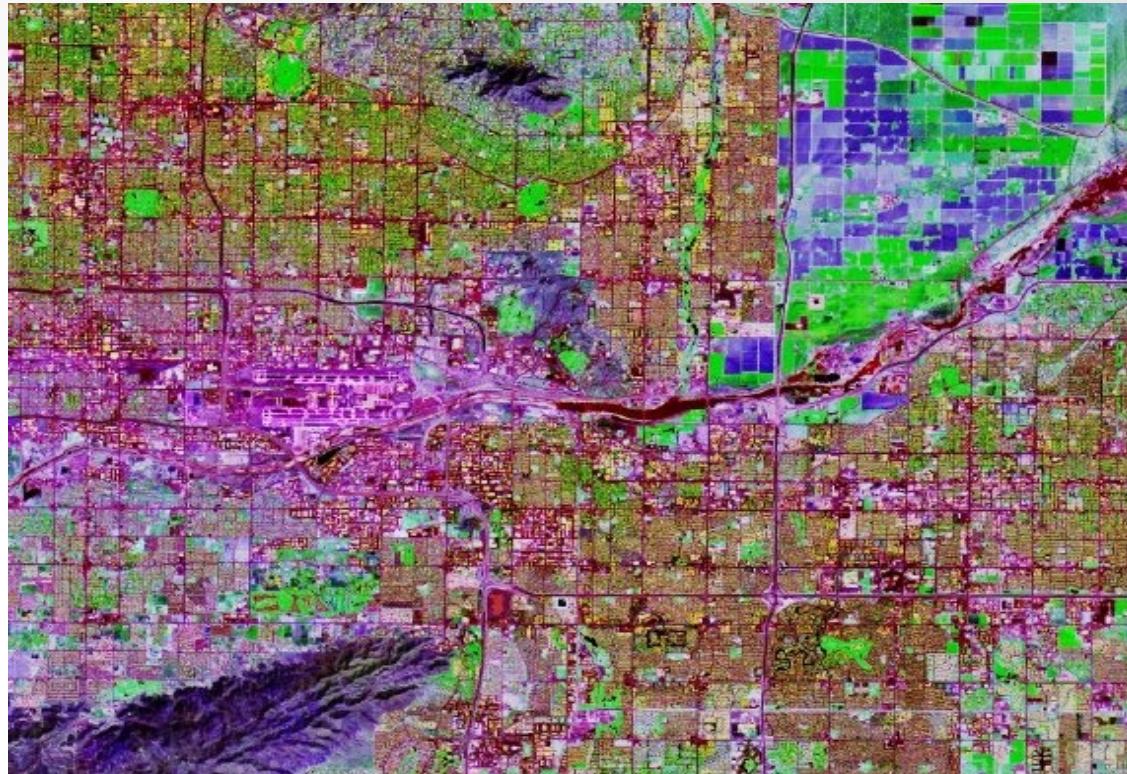
Climatology and Weather Prediction



Hurricane tracking

Requires sorting and processing millions of bytes per second

Global Positioning System (GPS) Data Processing



GPS location identification

*There are 31 active satellites orbiting at 20,000 km above the earth.
A GPS receiver calculates its 2D position (latitude and longitude)
by receiving signals from 3 satellites through a process called
triangulation.*

Geographical Location Data

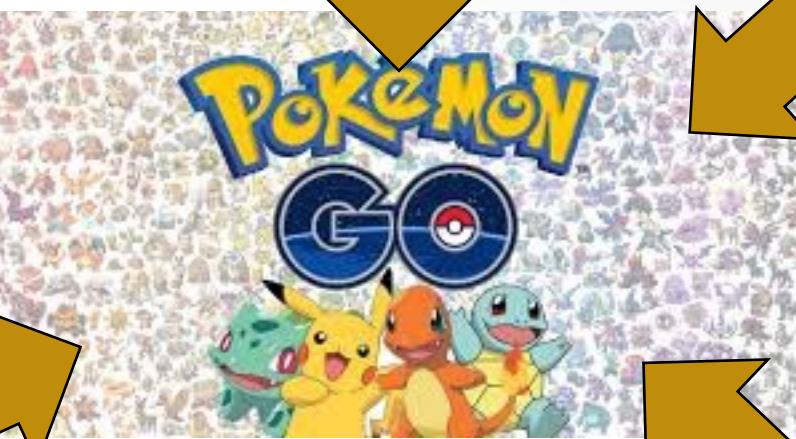
Cellphone GPS Tracking



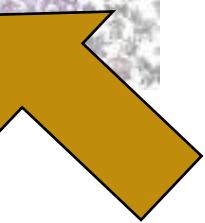
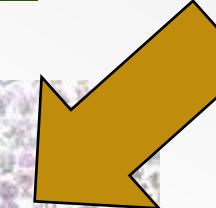
Pokemon Database



Augmented Reality and 3D Rendering



*Data Processing
and Algorithms in Pokemon GO*



Predictive Algorithms

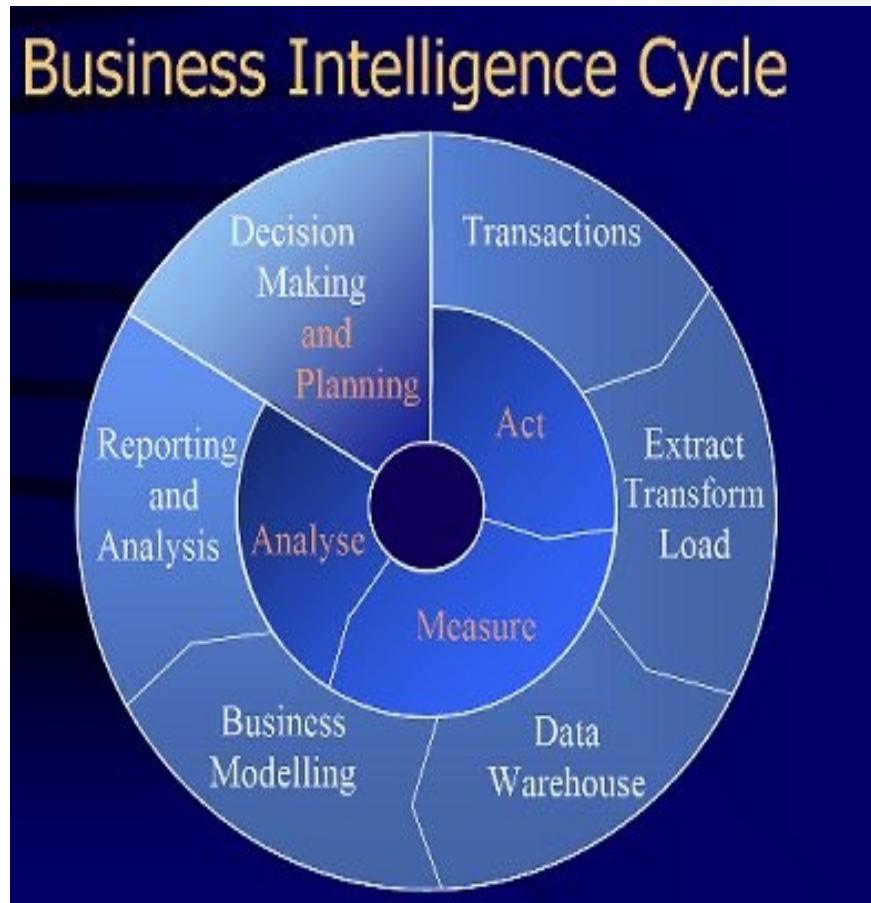


Nature and Data Structures: Did you know the relationship between a sunflower and Fibonacci numbers?



- Fibonacci series:
- 0,1,1,2,3,5,8,13,21,34,55,89...
- *The number of spirals in each direction is a Fibonacci number: 21 and 34, or 34 and 55, or 55 and 89, etc!*
- *The number of petals in a sunflower is also Fibonacci: 3, 5, 8, 13, 21, 34 or 55!*

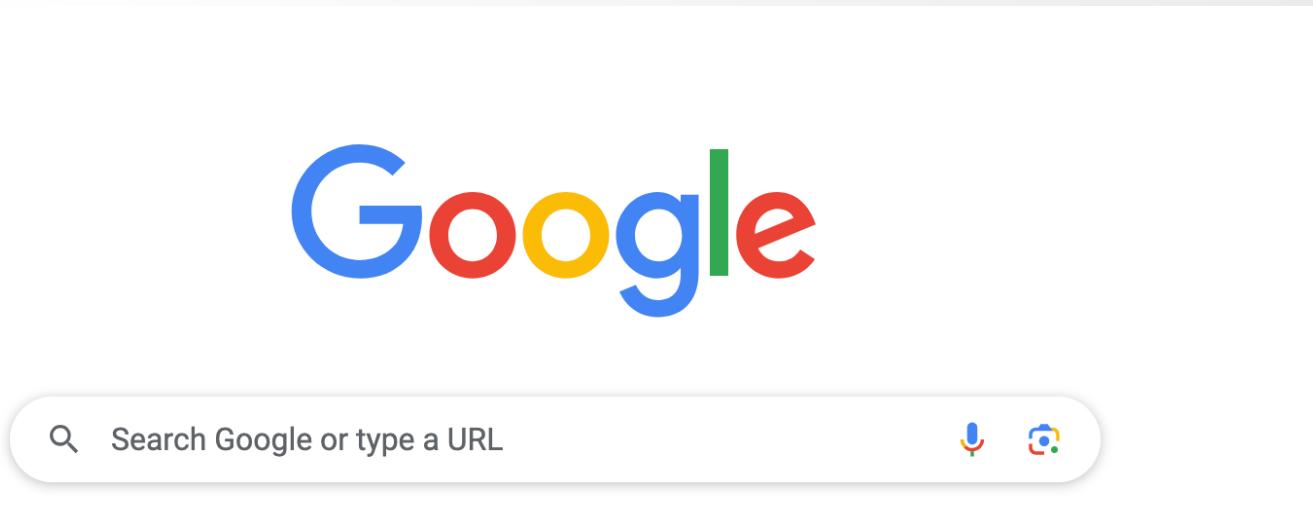
Business Intelligence



Data mining, online data analytics, querying and reporting

- Decision making – which products or services?
- Operations – streamline processes
- Forecasting and Predictive Analytics – future trends, demand, sales

How does a Google search work?



Several smart algorithms
including crawling, indexing, keyword analysis,
and a proprietary “Page Rank” algorithm.

What is the secret sauce in ChatGPT?



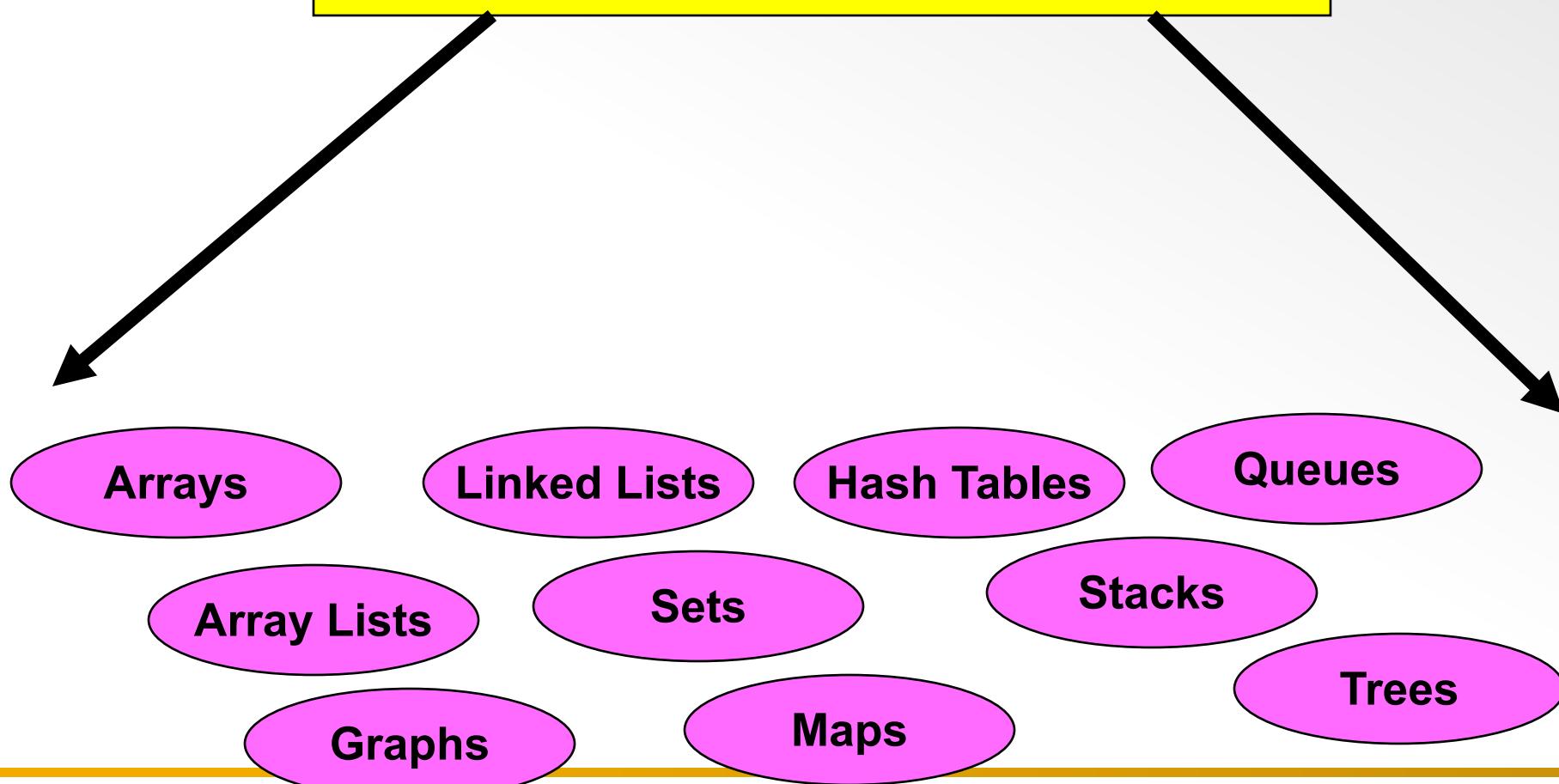
ChatGPT says:

ChatGPT doesn't use traditional search algorithms like indexing or crawling. Instead, it generates responses based on patterns learned from a very large corpus of text from the Internet, which allows it to learn grammar, facts and reasoning abilities.

COURSE OBJECTIVE

To provide the foundational knowledge in
Data structures and Algorithms
with implementation using the
object-oriented approach.

Data Structures
“Collections of related data items”



Choosing the right data structure for a given problem is the key to efficient software design.

Algorithms

- Methodology of solving the problem.
- Plethora of **standard algorithms** exist:
 - Searching
 - Linear, Binary, Depth first, Breadth first, Exhaustive, Greedy, ...
 - Sorting
 - Bubble sort, Insertion sort, Quick sort, Merge sort, Radix sort, Heap sort, Topological sort,
 - **Shortest path algorithms, Graph coloring, Traveling Salesperson, ...**

Choosing the right algorithm for the data structure is the next important key to efficient software design.

Complexity of algorithms can make a big difference!

- For example,
 - Sorting a million records with insertion sort: **1 billion steps**
 - Sorting a million records with heap sort: **20 million steps**
- As another example,
 - Linear search of a million records: **1 million steps**
 - Binary search of a million records: **20 steps**

Our approach to learning data structures and algorithms

For each data structure

- Understand its principles and concepts
- Understand its properties
- Understand its applications
- Design **algorithms** on it
- Analyze the **complexity** of the algorithm
- Implement it
- Apply it

Outcome

- ➔ *Teach you the essentials of data structures and algorithms.*
- ➔ *Easily digestible modules with lots of examples.*
- ➔ *Make it challenging but fun.*
- ➔ *Bridge the gap from first year CS.*
- ➔ *Ease the transition to third year.*
- ➔ *Prepare you for Co-op work term and employment.*

Course Logistics

Grading Scheme

Assignments :	20%
Labs:	10%
Short pop quizzes during lectures:	10%
Test 1:	15%
Test 2:	15%
Final exam:	30%

Assignments (Total: 20%)

Each assignment is **worth 4%** of your grade.

There will be six assignments, one of which is optional.

It means that the **best five** out of six assignments will be taken for grading.
You can choose to do all six assignments, or simply aim to do the five out of six.

Time for completion of each assignment is approx. 2 weeks.

Deadline is **11.59 PM** on the day of submission.

Up to **5 hours** grace time without late penalty.

Late penalty: 10% per day up to five days.

Labs (Total: 10%)

There will be approximately 10-11 labs through the term, one session per week.

You have the option of dropping two labs.

Each lab will be posted on Sunday on Brightspace.

Attend the synchronous lab session for which you have signed up.

Try to complete the lab during the lab session with TA help.

Labs will be due on Saturday at 11.59 PM of that week.

Again, 5 hours grace time without late penalty.

Late penalty: 10% per day up to two days.

Pop Quizzes (Total: 10%)

Short quizzes (approx. 10 minutes) occasionally given during lecture time.

Mainly for reflection, quick understanding of concepts, and class engagement.

Pop quizzes will be on Brightspace.

Pop quizzes are open book.

There will be 8 pop quizzes. You have the option to dropping two.

Tests and Final Exam

Most important assessment components!

They are closed book, in person and proctored.

They will be on Brightspace with Respondus Lockdown Browser.

Students with Accommodations: At the Accessibility Centre on Brightspace with Respondus Lockdown Browser.

Tests and Final Exam

Most important assessment components!

They are closed book, in person and proctored.

They will be on Brightspace with Respondus Lockdown Browser.

Students with Accommodations: At the Accessibility Centre on Brightspace with Respondus Lockdown Browser.

Test 1 is worth 15% of the overall grade.

Test 2 is worth 15% of the overall grade.

Final exam is worth 30% of the overall grade.

To pass the course, you must score a minimum of 50% out of the total weightage for the tests and the final exam. That is, your minimum mark for passing the course is 30 out of 60.

Student Declaration of Absence – Not required

Option to drop one assignment.

Option to drop two labs.

Option to drop two pop quizzes.

5 Hours grace time before late penalty kicks in.

Responsible Use of ChatGPT and Generative AI Tools

- You may use ChatGPT and other generative artificial intelligence (AI) tools in a responsible manner .
- Use them to enhance your understanding and supplement your learning outcomes.
- I will not use any ChatGPT checkers.
- You must acknowledge it if you used it for help with your assignments and labs.
- Tests and the Final Exam are closed book and in person, and will have no access to such tools.
-

Course Calendar

Sept. 5:	First lecture
Sept. 19:	Last day to register
Oct. 4:	Last day to drop course without a “W”
Oct. 11:	TEST 1 (6 – 7.30 PM)
Nov. 2:	Last day to drop course with a “W”
Nov. 8:	TEST 2 (6 – 7.30 PM)
Nov. 13 – 17:	Fall Study Break
Dec. 1:	Last Lecture
Dec. 8 - 19:	Final Exam period

References

There is no required textbook for the course.

I will be providing detailed course notes – Skeletal and complete.

In addition, I will be posting resources for extra reading.

Some references you may find useful:

1. *Data Structures Outside In with Java* by Sesh Venugopal, Pearson Prentice Hall Publishers ISBN 0-13-198619-8
2. *Data Structures and Problem Solving using JAVA* by Mark Allen Weiss, Addison Wesley, ISBN 0-201-74835-5
3. *Introduction to Java Programming (Comprehensive Version)* by Y.Daniel Liang, Pearson, ISBN 0-13-376131-2

Things to monitor

BRIGHTSPACE – One Stop Shop!

All course information will be posted on Brightspace.

Assignment and lab submission will be on Brightspace.

Pop quizzes, tests and final exams will be on Brightspace.

Important announcements will also be sent to your Dal email.

Office Hours for Help

Mondays 12 noon to 3 PM, Room 319, Goldberg CS Building.

I will also stay back after every lecture to answer your questions.

Brightspace course page is live.
Syllabus has been posted.
Please check!