DATA STRUCTURE AND ALGORITHM

Dr. Khine Thin Zar
Professor
Computer Engineering and Information Technology Dept.
Yangon Technological University

Lecture 14

Applications

Outlines of Class (Lecture 14)

- IP Router Tables
- Multi-Dimensional Packet Classification
- Web Information Retrieval
- Layout Data Structures
- Computer Graphics
- Geographic Information Systems
- Collision Detection
- Image Data Structures
- Data Mining

IP Router Tables

- Internet router classifies incoming packets into flows utilizing information contained in packet headers and a table of (classification) rules.
- The single field in the filter is the destination field and that the action is the next hop for the packet.
- Equivalent to the destination-based packet forwarding problem.

Multi-Dimensional Packet Classification

- Basic search algorithms, geometric algorithms, heuristic algorithms, or hardware-specific search algorithms
- Routers: provide different qualities of service to different applications such as admission control, resource reservation, per-flow queueing, and fair scheduling.
- Classifier: a collection of rules
- How packet classification can be used by an ISP to provide different services.

Multi-Dimensional Packet Classification (Cont.)

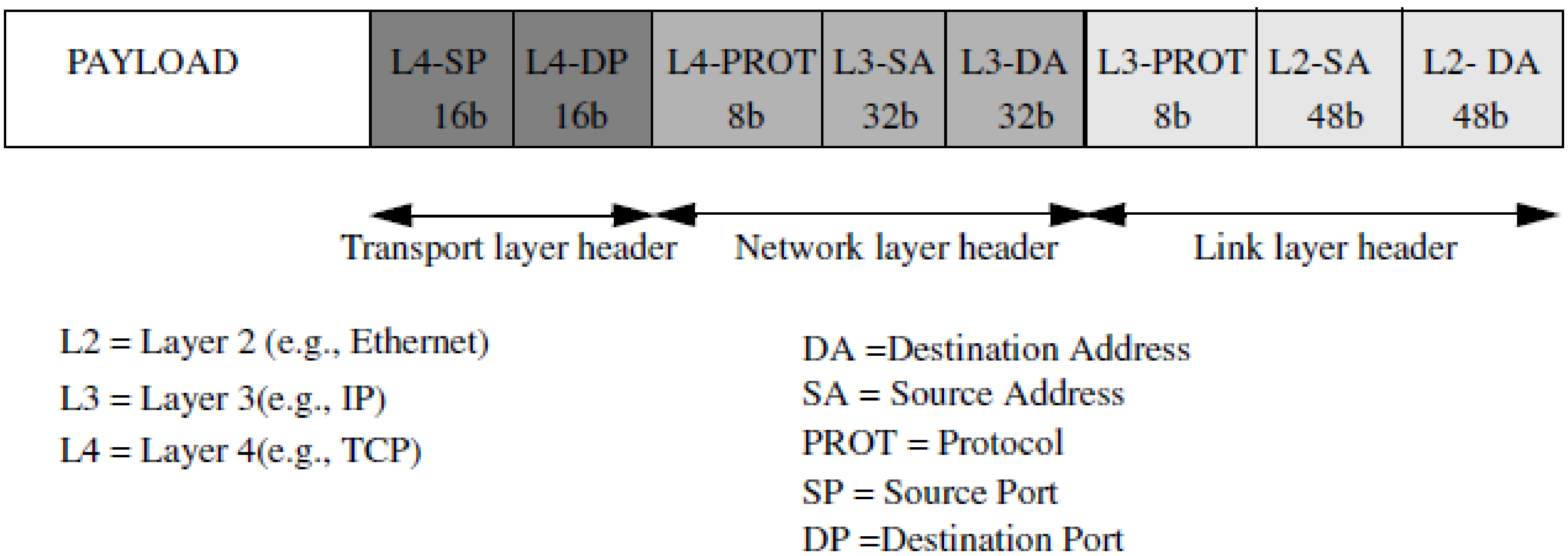

Figure: This figure shows some of the header fields (and their widths) that might be used for classifying the packet.

Performance Metrics for Classification Algorithms

- Search speed
- Low storage requirements
- Ability to handle large real-life classifiers.
- Fast updates
- Scalability in the number of header fields used for classification.
- Flexibility in specification

Taxonomy of Classification Algorithms

- Basic Data Structures
- Geometric Algorithms
- Heuristics
- Hardware-Based Algorithms

Web Information Retrieval

- Thousands of queries per second over a collection of billions of web pages with a sub-second average response time
- To find all documents that are relevant to a user query and return them in decreasing order of relevance.
- Web search engines usually return just the top 10 results and retrieve more results only upon request.
- To detect near-duplicate web pages

Inverted Indices

- Given a user query consisting of terms t1, . . . , tn,
- To find all documents relevant to the query.
- Uses the inverted (file) index data structure
- A data structure containing every word t that appears in at least one document together with a pointer p_t for each word and potentially a count f_t of the number of documents containing the word; and
- An inverted list for every term t that consists of all the documents containing t and is pointed to by p_t.
- To store the sorted list of words in a search tree whose leaves are connected by a linked list.

Fingerprints

- Short strings that represent larger strings and have the following properties:
- If the fingerprints of two strings are different then the strings are guaranteed to be different.
- If the fingerprints of two strings are identical then there is only a small probability that the strings are different.
- Search engines can use them to quickly check whether a URL that is contained on a web page is already stored in the index.

Finding Near-Duplicate Documents

- Users of search engine dislike getting near-duplicate results on the same results page.
- The main reasons are

(1) local copies of public-domain pages or various databases and

(2) multiple visits of the same page by the crawler without detection of the replication.

- The set of fingerprints for a document is computed using the shingling technique

Layout Data Structures

- VLSI (Very Large Scale Integration) is a technology that has enabled the manufacture of large circuits in silicon.
- The purpose of VLSI design automation is to develop software that is used to design VLSI circuits.
- The purpose of a layout data structure is to store and manipulate the rectangles on each layer.
- High-level operations:
  - ✓ design-rule checking,
  - ✓ layout compaction, and
  - ✓ parasitic extraction.

Quad Trees and Variants

- Different flavors of quad-trees depending on the type of the data that are to be represented.
- Concerned with quad-trees for rectangles
- To recursively subdivide the two-dimensional
- Layout area into four "quads" until a stopping criterion is satisfied
- The resulting structure is represented by a tree with a node corresponding to each quad

Quad Trees and Variants (Cont.)

- Three strategies
  - ✓ SMALLEST: Store a rectangle in the smallest quad (not necessarily a leaf quad) that contains it. Such a quad is guaranteed to exist since each rectangle must be contained in the root quad.
  - ✓ SINGLE: Store a rectangle in precisely one of the leaf quads that it intersects.
  - ✓ MULTIPLE: Store a rectangle in all of the leaf quads that it intersects.

Quad Trees and Variants (Cont.)

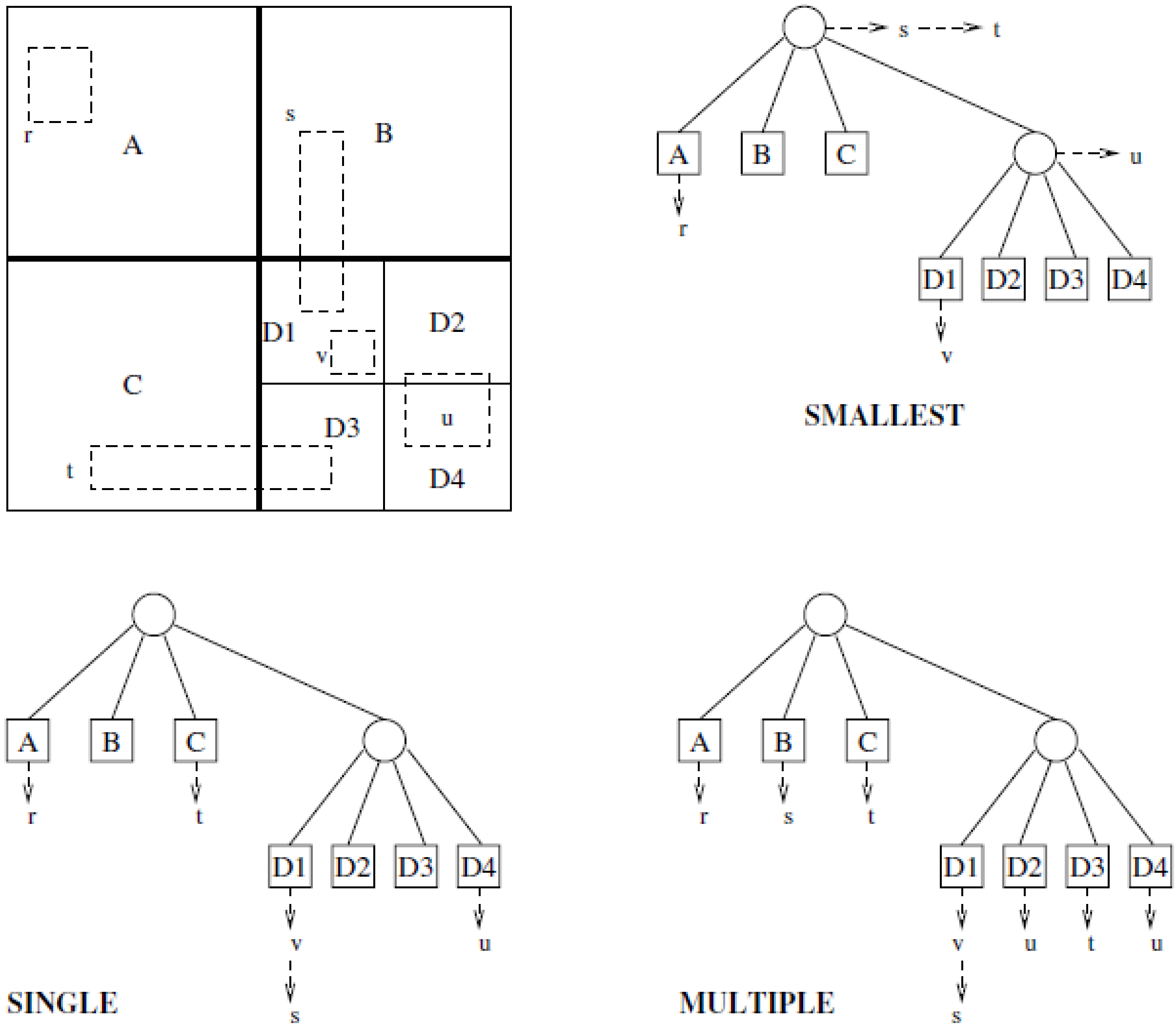

Figure: Quadtree variations

Floorplan Representation in VLSI

- Two main data models to represent floorplans:
- Graph-based
  - ✓ constraint graphs: basic representations
  - ✓ corner stitching: simplifies the constraint graph by recording only the four neighboring blocks to each block
  - ✓ twin binary tree: reduces the recorded information to only two neighbors of each block, and organizes the neighborhood relations in a pair of binary trees
  - ✓ O-tree: further simplification to the twin binary tree
- Placement based
  - ✓ sequence pair, bounded-sliceline grid: applied to general floorplan
  - ✓ corner block list: records only the relative position of adjacent blocks
  - ✓ slicing trees: for slicing floorplan

Statement of Floorplanning Problem

- Inputs:

    1. the net-lists of the sub-circuits;

    2. the area estimation of blocks and, if any, the aspect ratio constraints on the blocks;

    3. the target area and shape of the entire chip.

- Outputs:

    1. the shapes and positions of blocks;

    2. the pin positions on the blocks.

- The objective functions involve: the chip area, the total wire length and, the performances.

Statement of Floorplanning Problem (Cont.)

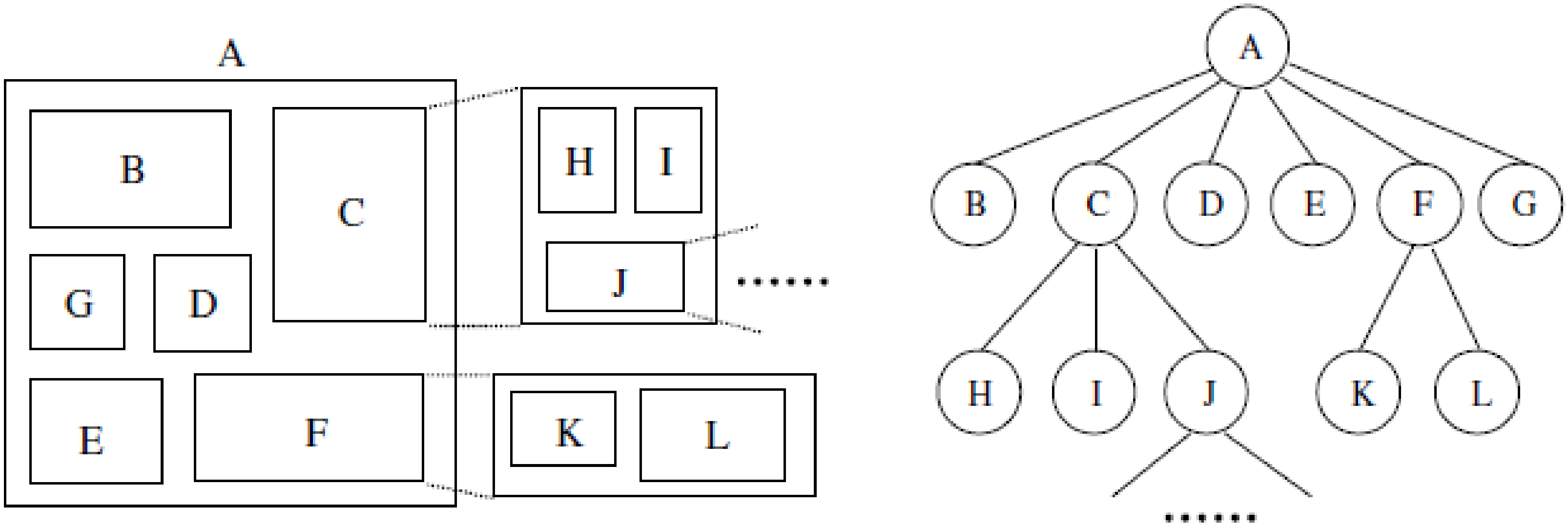

Figure: Hierarchical structure of blocks.

Computer Graphics

- The computer graphics industry depends on the efficient synergy of hardware and software
- As hardware capabilities grow, the potential for new feasible uses for computer graphics emerge
- The needs for graphics applications grow and change
- The way graphics primitives are represented, or stored in computer memory
- Use of many standard, stable data structures, algorithms, and models.

Computer Graphics (Cont.)

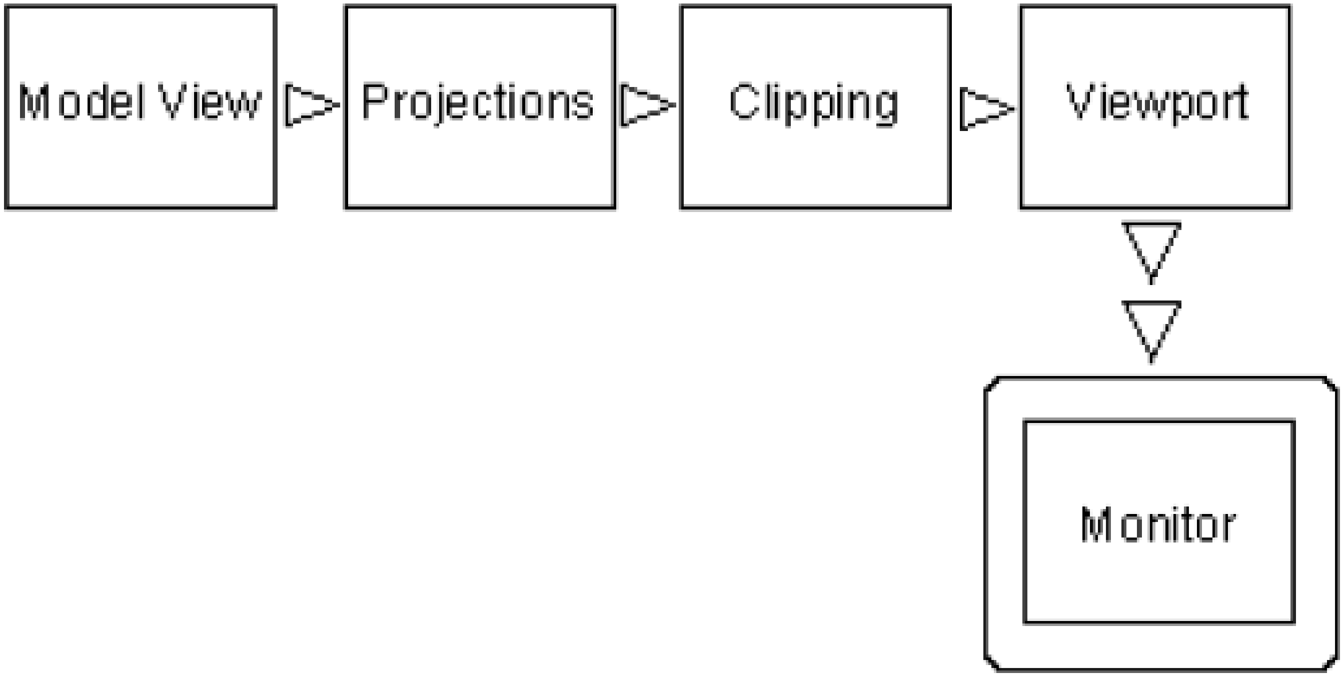

Figure: Graphics pipeline

Basic Applications

- Meshes
- CAD/CAM Drawings
- Fonts
- Bitmaps
- Texture Mapping

Geographic Information Systems

- The purpose of maintaining, analyzing and visualizing spatial data that represent geographic objects
- Geodesic measurement points, boundary lines between adjacent pieces of land, lakes or recreational park regions
- The data structures and algorithms aspects of GISs

Applications of Geographic Information

- Map Overlay
- Map Labeling
- Cartographic Generalization
- Road Maps
- Spatiotemporal Data
- Data Mining

Models, Toolboxes and Systems for Geographic Information

- The core functionality of a GIS related to manipulating vector data
  - ✓ Data Model: The spatial data model offered by the system is very important to a user since it provides the geometric data types and the operations.
  - ✓ Spatial indexing: Spatial index structures are important for efficiently supporting the most important spatial queries.
  - ✓ Spatial Join: Since spatial joins are the most important operation for combining different maps, system performance depends on the efficiency of the underlying algorithm.

Collision Detection

- ❑ Refers to checking the relative configuration of two or more objects
- ❑ To automatically report a geometric contact when it is about to occur or has actually occurred
- ❑ Polygonal objects, spline or algebraic surfaces, deformable models, etc (the objects may be static or dynamic)
- ❑ Arises in different applications
  - ✓ Physically-based Modeling and Dynamic Simulation
  - ✓ Motion Planning
  - ✓ Virtual Environments and Walkthroughs
  - ✓ Haptic Rendering

Architecture for Multi-body Collision Detection

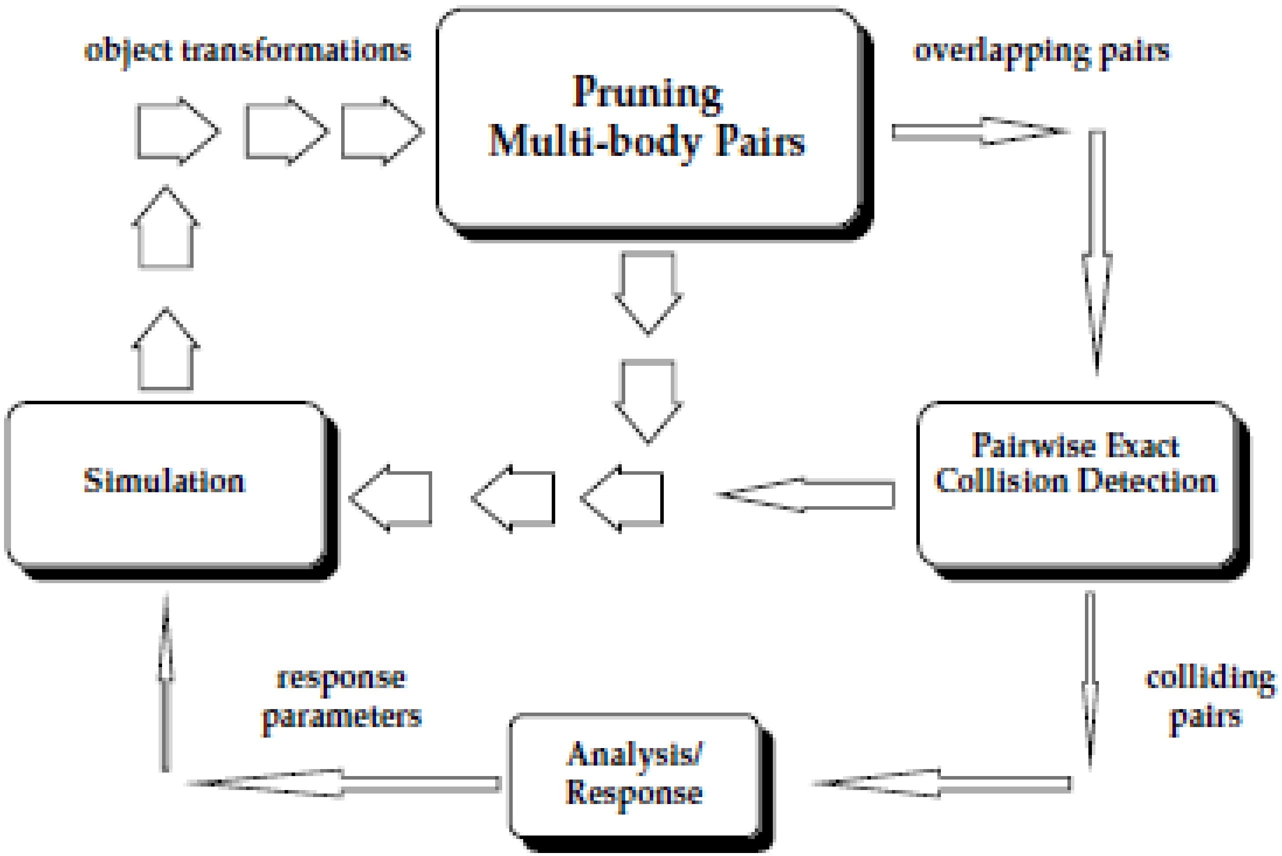

Figure: Architecture for multiple body collision detection algorithm

Image Data Structures

- Visual information aids us in understanding our surroundings better
- Used for digitally interpreting images
- Three main steps

1. Image acquisition: Obtaining the image by scanning it or by capturing it through some sensors.
2. Image manipulation/analysis: Enhancing and/or compressing the image for its transfer or storage.
3. Display of the processed image.

Image Data Structures (Cont.)

- Image processing has been classified into two levels
- Low-level image processing:
  - needs little information about the content or the semantics of the image
  - concerned with retrieving low-level descriptions of the image and processing them
  - include matrix representation of the actual image
  - E.g., Image calibration and image enhancement
- High-level image processing:
  - concerned with segmenting an image into objects or regions.
  - The high-level data include features of the image

Data Mining

- an explosive growth in the amounts of data collected, stored, and disseminated by various organizations

(1) the large volumes of pointof- sale data amassed at the checkout counters of grocery stores,

(2) the continuous streams of satellite images produced by Earth-observing satellites, and

(3) the avalanche of data logged by network monitoring software.

- how to extract useful information from massive data sets?

Data Mining (Cont.)

- Data mining is an integral part of another process, Knowledge Discovery in Databases (or KDD)
- The overall process of turning raw data into interesting knowledge
- Consists of a series of transformation steps
  - ✓ Data preprocessing: to convert data into the right format for subsequent analysis
  - ✓ Data mining: feature selection and construction
  - ✓ Postprocessing: all additional operations performed to make the data mining results more accessible and easier to interpret.

Data Mining Tasks and Techniques

- Data mining tasks are divided into two major categories:
  - ✓ Predictive: to use the values of some variables to predict the values of other variables
  - ✓ Descriptive: to find human-interpretable patterns that describe the underlying relationships in the data
- Data mining techniques
  - ✓ Predictive modeling
  - ✓ Association rule mining
  - ✓ Cluster analysis
  - ✓ Anomaly detection

End of the Course…

Ref: Handbook of data structures and applications / edited by Dinesh P. Mehta and Sartaj Sahni

Thank you!