高级数据结构

Advanced Data Structure

1. Introduction

靳晓明 Jin Xiaoming

School of Software, Tsinghua University

Agenda

- Why ADS
- Major issues in ADS
- Course information

Fundamental structures meets real applications

- Very large databases
- **Challenge 1:** How to access it efficiently?

- Usually, we want deal with subjective notions
- **Challenge 2**: How to handle the multimedia data based on their content?

What's ADS

- Data structure and algorithms
  - Simple data objects
  - Mostly main-memory based data
  - Fundamental data structure and algorithm

- **ADS: Data structures and algorithms concerning**
  - Complex data objects
  - Huge data amount
  - Real applications

"Complex data objects"

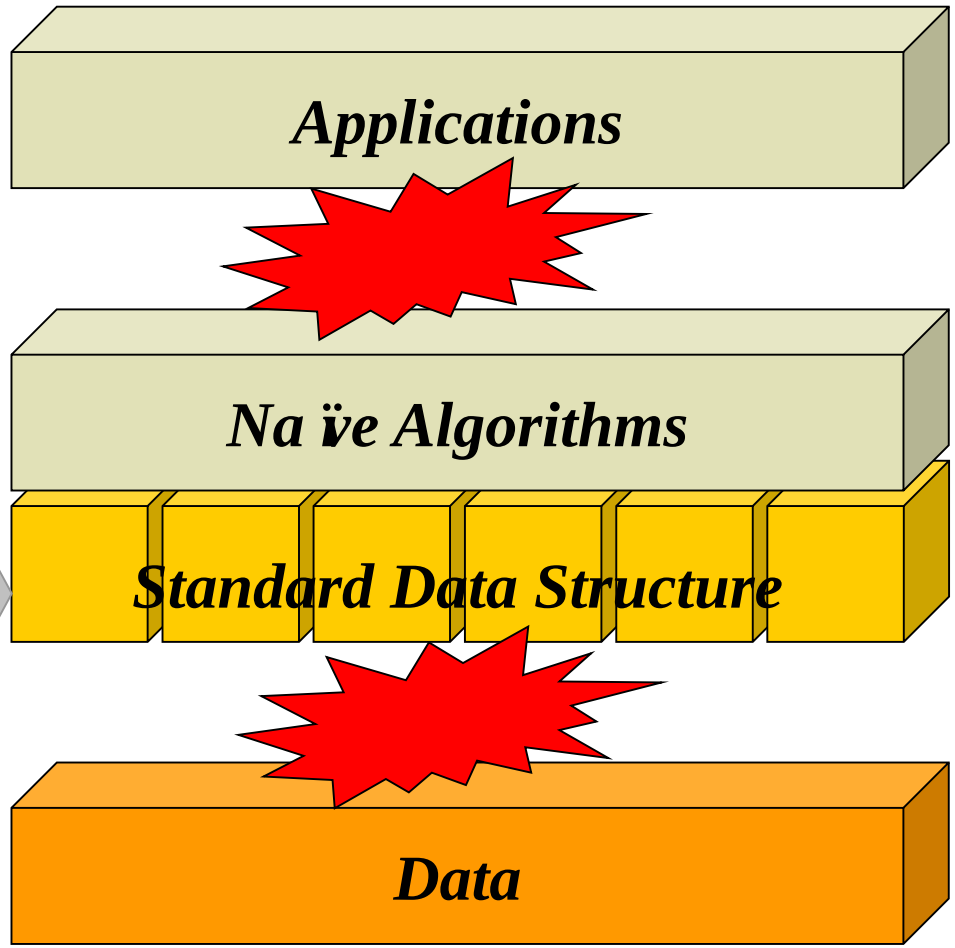
- Real variables
- Points
  - Vector that contain two or three values (x,y) (x,y,z)
- Spatial objects
  - A real value vector (x1,y1,x2,y2) (x,y,r)
- Text
  - Sequence of symbols (S1S2…Sn)
- Image
  - A matrix of value or 3-tuple
- Time series
  - Sequence of real values (x1,x2,…,xn)
- Video
  - A sequence of images (I1,I2,…,In)

"Complex accessing task" – data mining

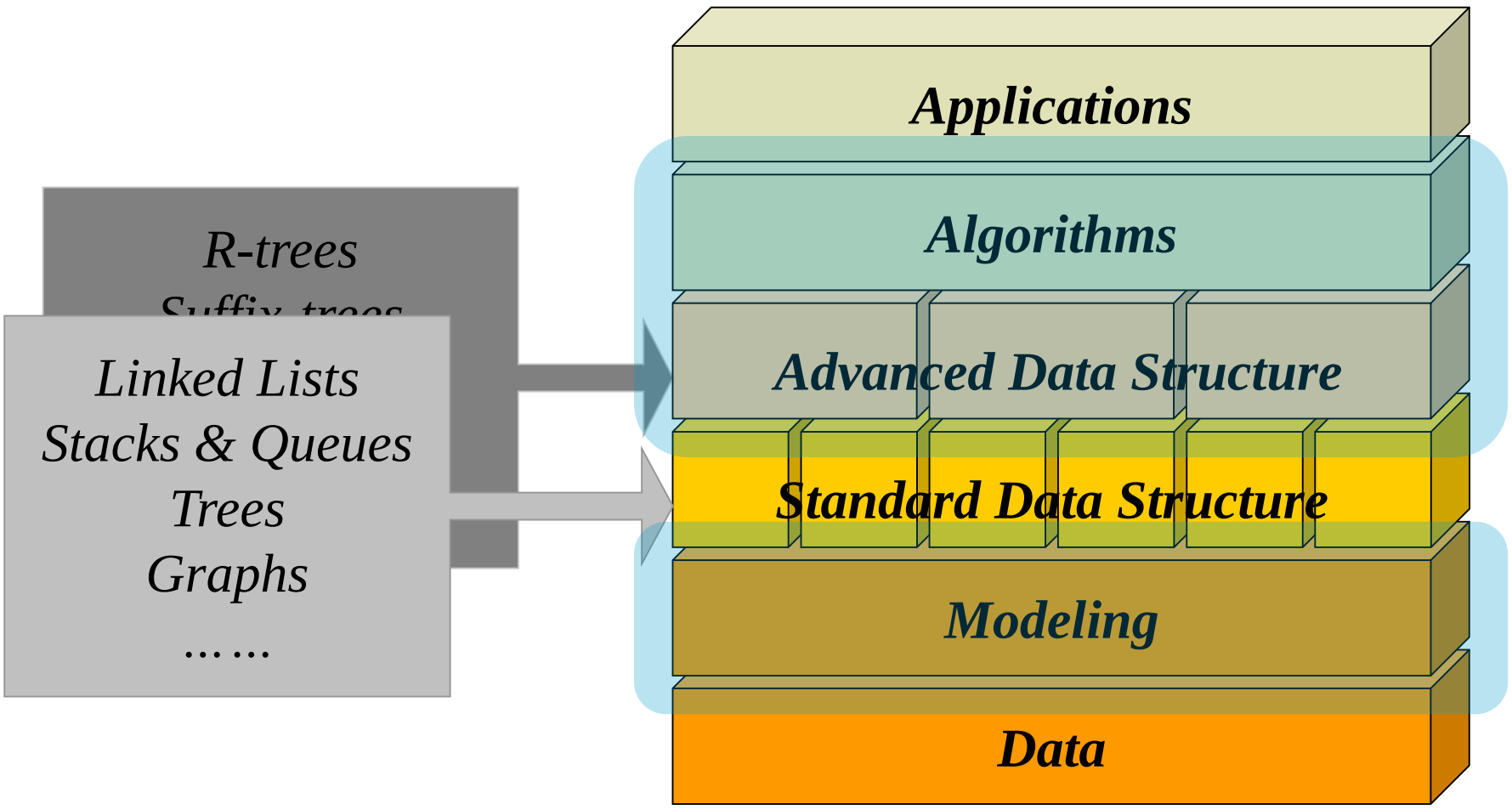
Where is ADS?

How to use data? effectively?
and efficiently?
How to fill the gap from data to applications?

Linked Lists
Stacks & Queues
Trees
Graphs
……



Where is ADS?


R-trees
Linked Lists
Stacks & Queues
Trees
Graphs
……
Applications
Algorithms
Advanced Data Structure
Standard Data Structure
Modeling
Data

Agenda

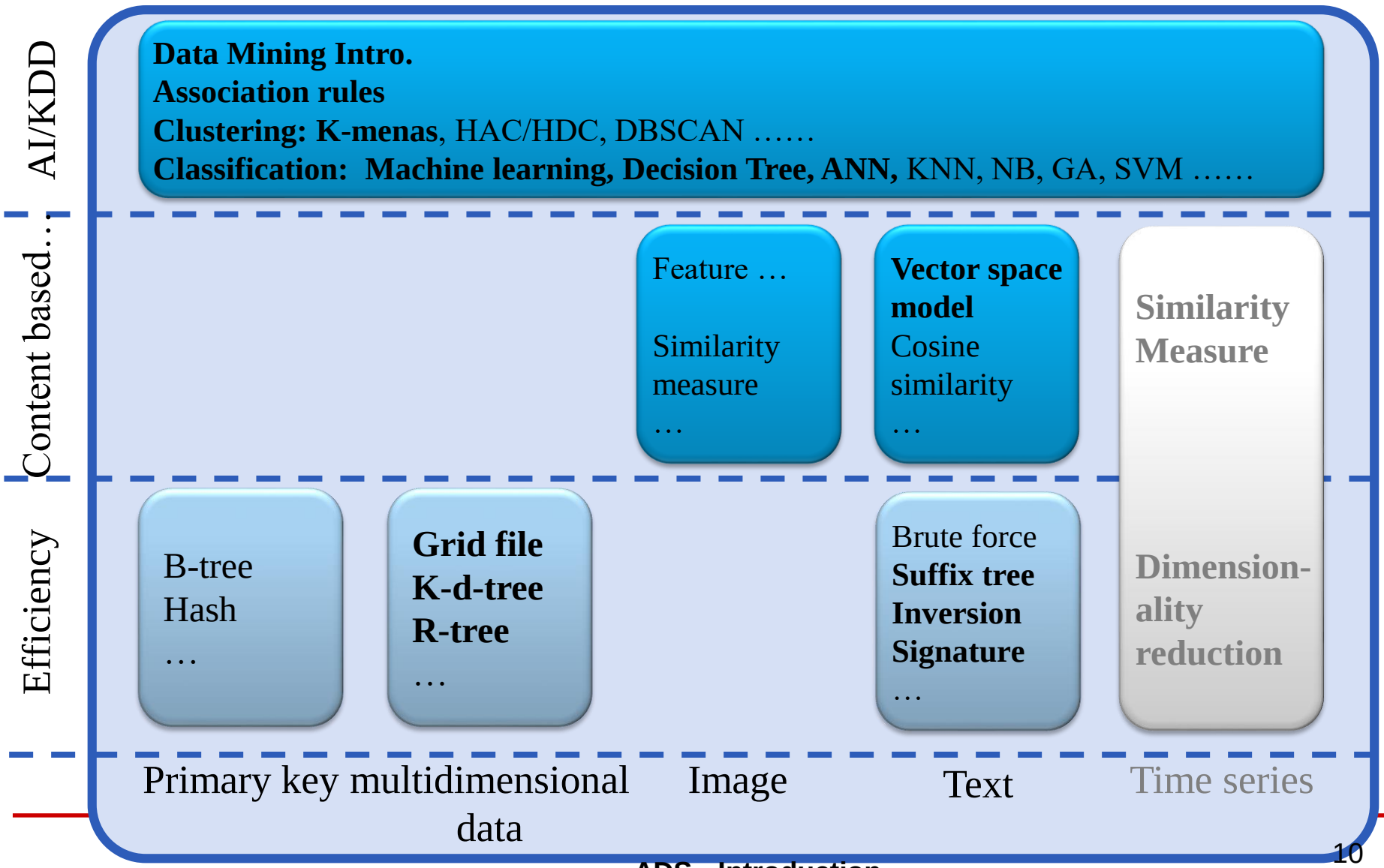
- Why ADS
- Major issues in ADS
- Course information

About this course - Purpose

- To present the key data structure, algorithms, theory, and application issues that form the core of data accessing

- To people
  - who may need to solve problems in storing and retrieving complex data objects
  - who do research in data mining, database, information retrieval…

Advanced Data Structure – The Big Picture



Related topics/Courses

- Data Structure
- Database
  - Advanced database techniques
  - Multimedia database
- Information Retrieve, Computer Vision
- Data Mining, Pattern Recognition
- AI, Machine Learning
- Applications:
  - Internet, Bio-informatics, Mobile device, Sensor network,
  - Medical care, CAD/CAM , Business Intelligence, Games …

Agenda

- What’s ADS
- Major issues in ADS
- Course information

About this course - Grading

- **Discussion and in-class exercises 40%**
  - In-class exercises
  - Presentation and discussion
- **Experimental project 20%**, teamwork (≤ 3 members)
  - Report
- **Course project 40%**, teamwork (≤ 3 members)
  - Discussion, final report, demo and presentation

- **Policy on grading**:
  - Inter-team collaboration on project may be seen as cheating → immediate failure
  - Collaboration on exercises is cheating → immediate failure
  - Submitted report is worth full credit at the ending of class on the due date, it is worth half credit for the next class, and ZERO credit after that.

About this course – Text book

- **Textbook:**
  - **Primary textbook:**
    C. Faloutsos
    Searching Multimedia Databases by Content**,**
    Kluwer Academic Press

  - Recommended textbook:
    Modern information retrieval
    Machine Learning
    Data Mining: Concepts and Techniques

  - Selected papers from main-stream conferences
    SIGMOD, VLDB, ICDE
    SIGIR, WWW, ICCV
    KDD, ICML, IJCAI, AAAI
    TKDE, TOIS, TODS …

Thanks

Feedback welcome