

CME2002

Data Organization and Management (DOM)

Lecture #1

Dr. Adil Alpkoçak

2016 Spring



Course Objectives

“The main objectives of this course are to discuss and teach

1. **Data**, **information** and **Knowledge** triple, **organization** and **processing of files**, physical **characteristics of storage media**,
2. **file processing techniques** and **indexing structures** for fast access to data, information and knowledge.
3. Introduction to DBMS and SQL



Teaching Team

- ◆ Assoc.Prof.Dr. Adil ALPKOÇAK
alpkocak@cs.deu.edu.tr
- ◆ Res.Asst. Dr. Mete Uğur Akdoğan
- ◆ Res.Asst. Mansur Alp Toçoğlu
- ◆ Res.Asst. Mehmet Cengiz





6wy00dn

Google
Classroom

- ◆ This year **Google Classroom** will be used as a Course Management System, instead of Department's Moodle Server.
- ◆ Class enrollment code **6wy00dn**



Prerequisites

1. Students who wants to take this course must have taken and pass the course CME2001 Data structures and Algorithms. Some of the techniques will be discussed based on the fundamental data structure.
2. Knowledge in C programming language is highly requested. This is because some of the assignments requires coding in C programming language. A short tutorial will be given in lab sessions.



Weekly Schedule

- ◆ Lecture session
(Class D4)
 - ◆ Wednesday Afternoon

- ◆ Lab practice session
(Computer Labs)
 - ◆ Thursday Afternoon



Assignments

- ◆ During the course term, 3 assignments will be given.
- ◆ All the assignments must be meticulously documented and must be returned in time.
- ◆ No late assignments will be accepted!
- ◆ However, assignments must be completed and submitted to take final exam.



Academic dishonesty

- ◆ You can discuss the implementations and details of the assignments with your friends but always remember that assignments must be your **individual** work, otherwise stated.
- ◆ Cheating is strictly prohibited !!!
- ◆ If any cheating occurs, both of the works/ assignments will be graded with **zero**. If it repeats department administration will be informed.



Grading Policy

Midterm exams	25%
3 Assignments	25%
Final Exam	50%



Date	Date	Topics
1	Feb, 22-23	Welcome, DIKW, C intro
2	Mar, 01-02	XML, DTD, XSD
3	Mar, 08-09	XSL and XQuery
4	Mar, 15-16	XML Applications
5	Mar, 22-23	Storage devices & Basic File Ops.
6	Mar, 29-20	Midterm Exam
7	Apr, 05-06	Analytical modelling & Simple Indexing
8	Apr, 12-13	B+tree
9	Apr, 19-20	Multidimensional Indexing
10	Apr, 26-27	External Sorting
11	May, 03-04	ER Model and Normalization
12	May, 10-11	Midterm Exam
13	May, 17-18	DBMS and SQL
14	May, 24-25	Structured Query Language



Motivation

- Most computers are used for **data processing**, as a big growth area in the “information age”
- **Data** processing from a computer science perspective:
 - ◆ Storage of data
 - ◆ Organization of data
 - ◆ Access to data
 - ◆ Processing of data



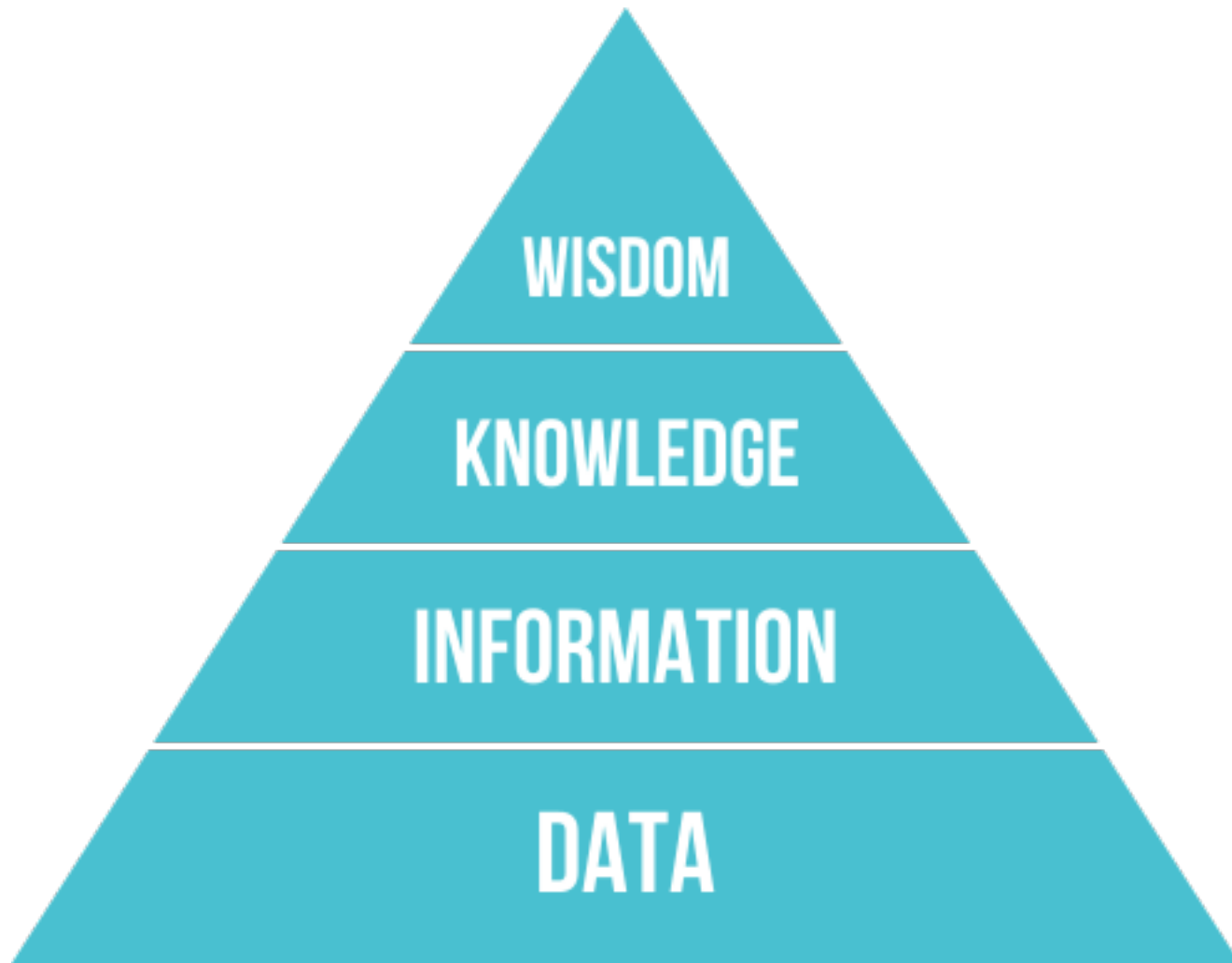
How to succeed in this course

- ◆ Attend and participate classes. Ask question whenever you find something unclear.
- ◆ Don't miss the lab session !!
- ◆ Do the assignments, don't wait till the last night to start before submission!!
- ◆ Ask help whenever you studied enough and don't understand something.
- ◆ Study well before exams, review previous year questions.

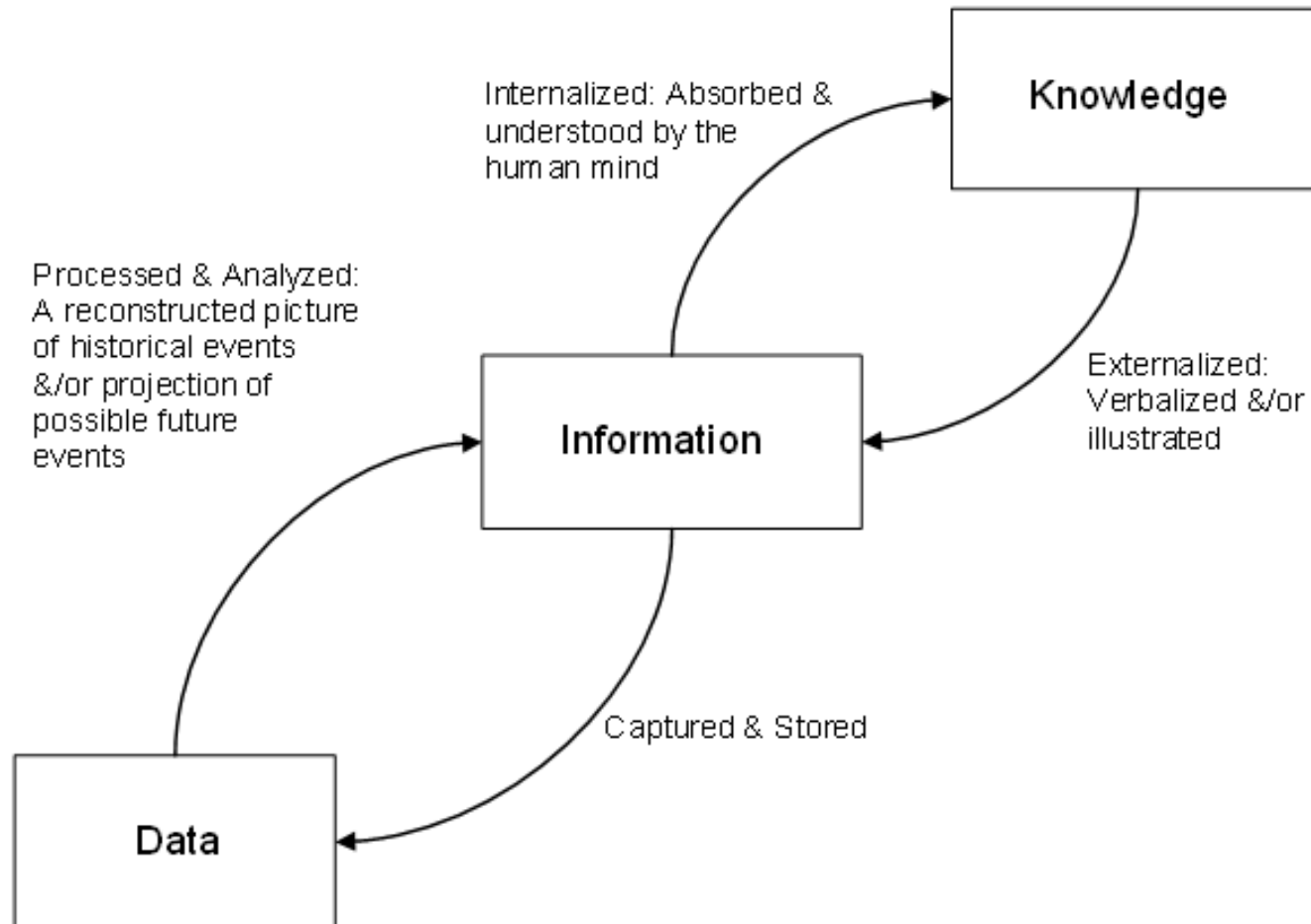


CME2002 Data Organization & Management





Relationships Amongst Knowledge, Information, and Data



Data

- ◆ Data **are** raw facts and figures that on their own have no meaning
- ◆ These can be any alphanumeric characters i.e. text, numbers, symbols



Data Examples

- ◆ Yes, Yes, No, Yes, No, Yes, No, Yes
- ◆ 42, 63, 96, 74, 56, 86
- ◆ 111192, 111234
- ◆ None of the above data sets have any meaning until they are given a **CONTEXT** and **PROCESSED** into a useable form



Data Into Information

- ◆ To achieve its aims the organisation will need to **process** data into information.
- ◆ Data needs to be turned into meaningful information and presented in its most useful format
- ◆ Data must be processed in a **context** in order to give it meaning



Information

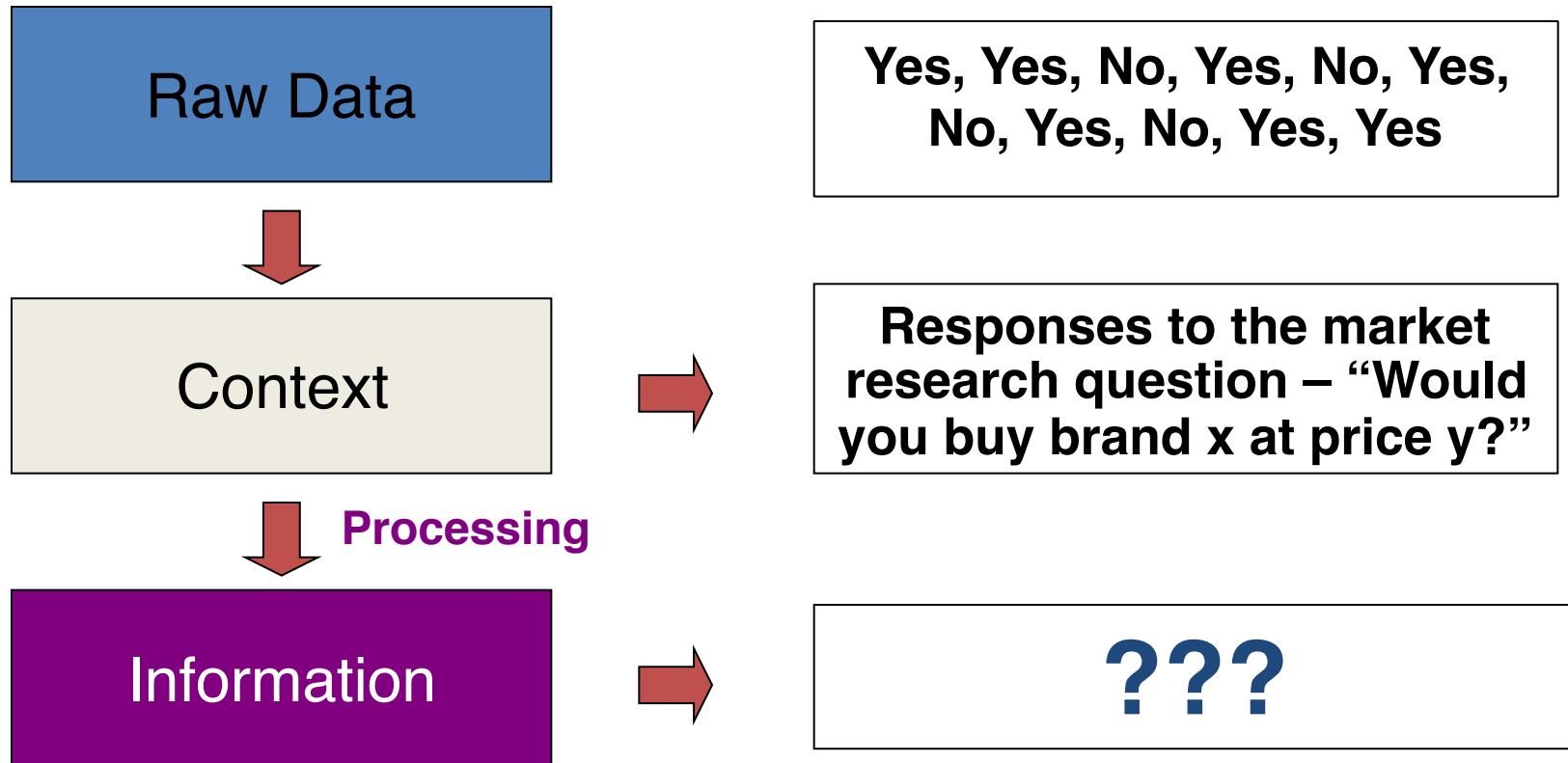
- ◆ Data that has been processed within a context to give it meaning

OR

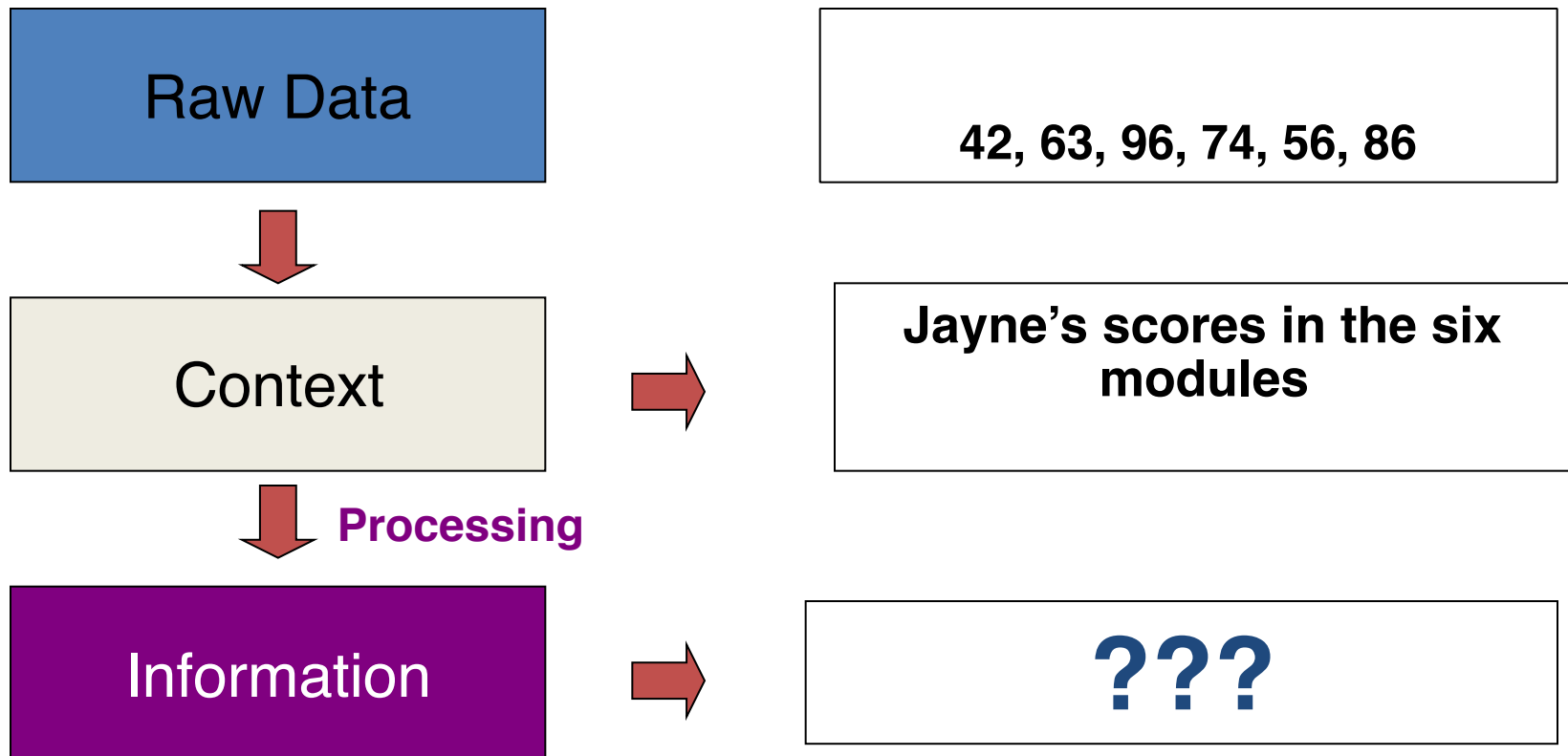
- ◆ Data that has been processed into a form that gives it meaning



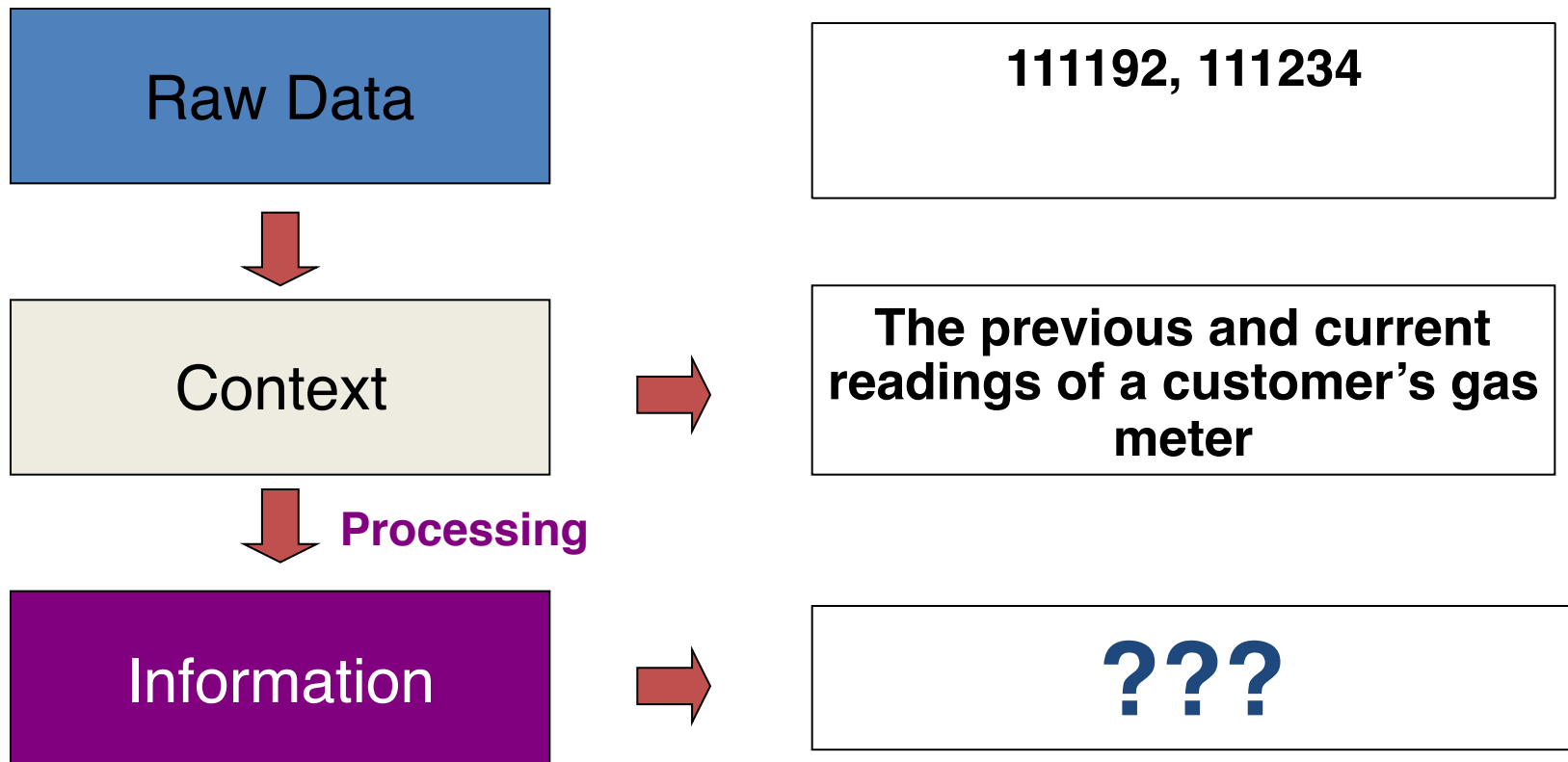
Example 1



Example 2



Example 3



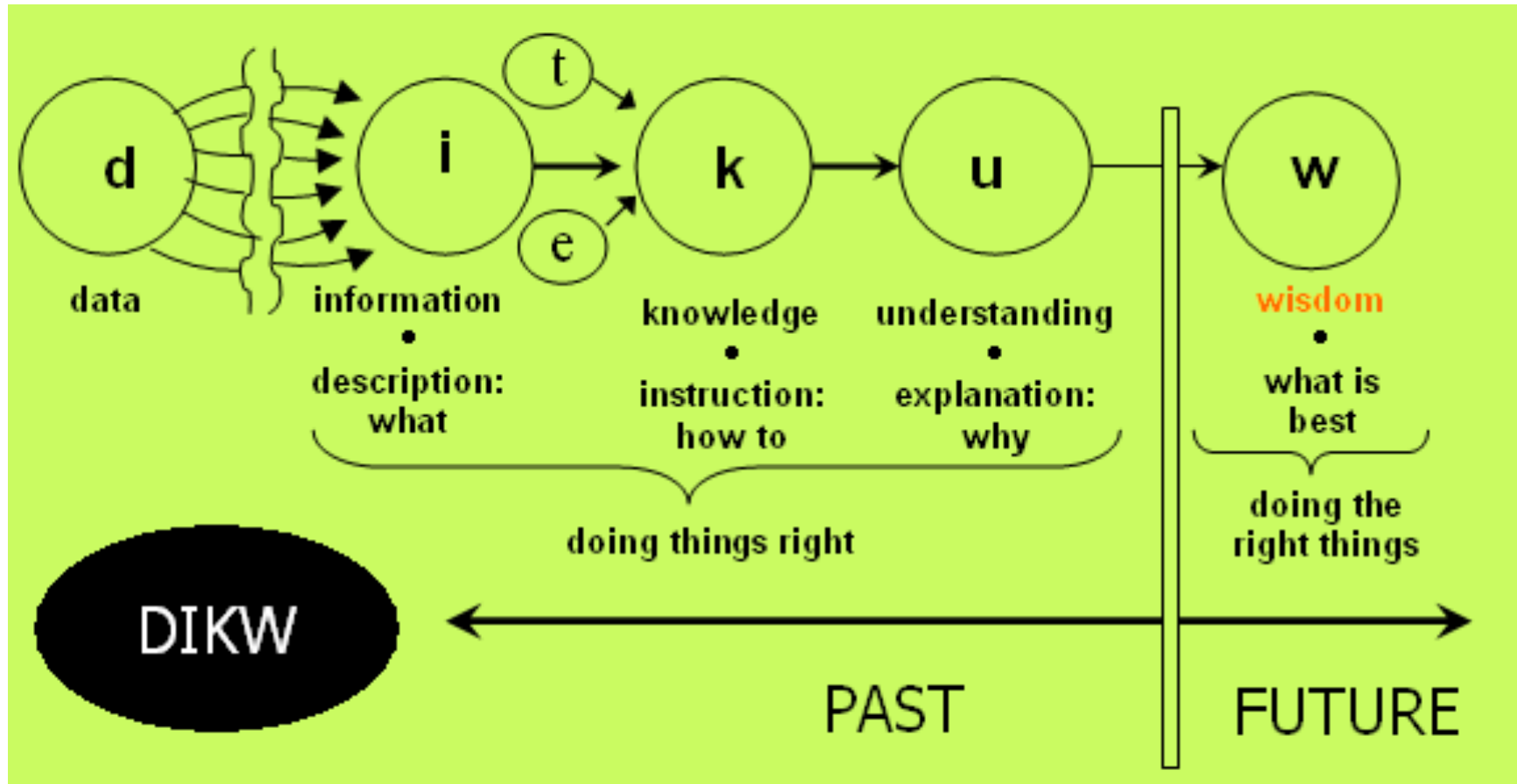
Knowledge

- ◆ Knowledge is the understanding of rules needed to interpret information

“...the capability of understanding the relationship between pieces of information and what to actually do with the information”

Debbie Jones – www.teach-ict.com





Knowledge

- Data and information deal with facts and figures
- Knowing what to do with them requires **knowledge**
- **Knowledge = information + rules**
- Rules tell us the likely effect of something
- For example: you are more likely to pass your A level IF you do your coursework and revise for your exam!



Knowledge Examples

- ◆ Using the 3 previous examples:
 - ◆ A Marketing Manager could use this information to decide whether or not to raise or lower price y
 - ◆ Jayne's teacher could analyse the results to determine whether it would be worth her re-sitting a module
 - ◆ Looking at the pattern of the customer's previous gas bills may identify that the figure is abnormally low and they are fiddling the gas meter!!!

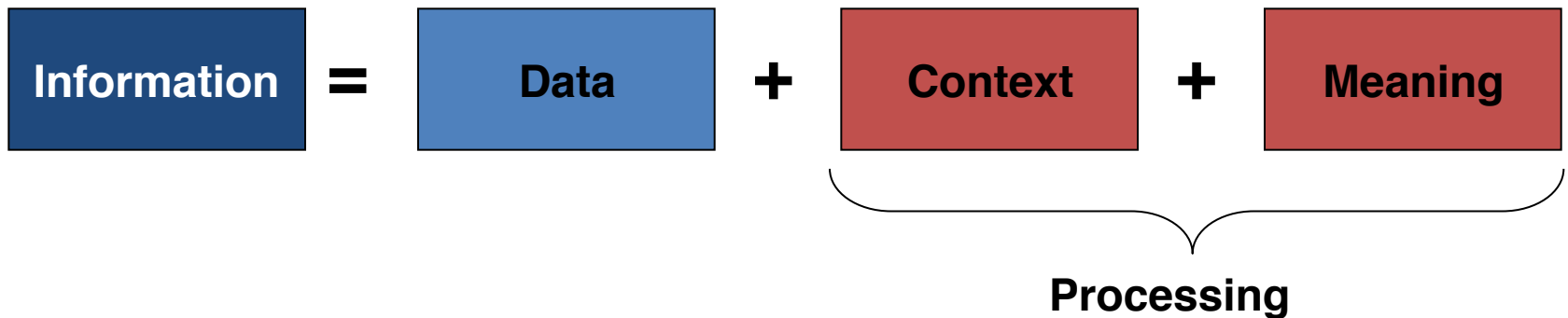


Knowledge Workers

- ◆ Knowledge workers have specialist knowledge that makes them “experts”
 - ◆ Based on formal and informal rules they have learned through training and experience
- ◆ Examples include doctors, managers, librarians, scientists...



DIK - Summary



Data – raw facts and figures

Information – data that has been processed (in a context) to give it meaning



Data Representation

- ◆ Bit
- ◆ Byte
- ◆ Numbers (signed, unsigned, floating point)
- ◆ Character
- ◆ Text
- ◆ Image
- ◆ Sound



Prefixes for bit and byte multiples

Value	Binary	Symbol	prefix
10^3	2^{10}	k	kilo
10^6	2^{20}	M	mega
10^9	2^{30}	G	giga
10^{12}	2^{40}	T	tera
10^{15}	2^{50}	P	peta
10^{18}	2^{60}	E	exa
10^{21}	2^{70}	Z	zetta
10^{24}	2^{80}	Y	yotta



Data Organization & Management

“Data is stored in files”



Data Structures vs File Structures

- ◆ Both involve

Representation of Data

+

Operations for accessing data

- ◆ Difference:

- ◆ **Data structures** deal with data in the main memory
- ◆ **File structures** deal with the data in the secondary storage



Goal of the File structures

◆ Performance

◆ Time

- ◆ Minimize the number of **hops** in order to get desired information
- ◆ Group related information so that we are likely to get everything we need with fewer hops.

◆ Memory

- ◆ Balance the memory size and the time

◆ How to improve performance

◆ Use the right file structure

- ◆ Understand the advantages disadvantages of alternative methods



Metrics used to measure efficiency and effectiveness of a File structure

- ◆ Time complexities,
- ◆ Space complexities,
- ◆ Simplicity,
- ◆ Reliability,
- ◆ Scalability,
- ◆ Programmability, and
- ◆ Maintainability.



Metrics used to measure efficiency and effectiveness of a File structure

- ◆ The file structures involve two domains: **hardware** and **software**.
 - ◆ Hardware - physical characteristics of the storage medium.
 - ◆ Software - data structures and algorithms to deal with these structures.
- ◆ Hardware + SW (data structures and algorithms) are used to predict the efficiency of file operations.



Basic File operations

- ◆ Search for a particular data in a file,
- ◆ Add a certain data item,
- ◆ Remove a certain item,
- ◆ Order the data items according to a certain criterion,
- ◆ Merge of files,
- ◆ Creation of new files from existing file(s).
- ◆ create, open, and close operations which have implications in the operating system.

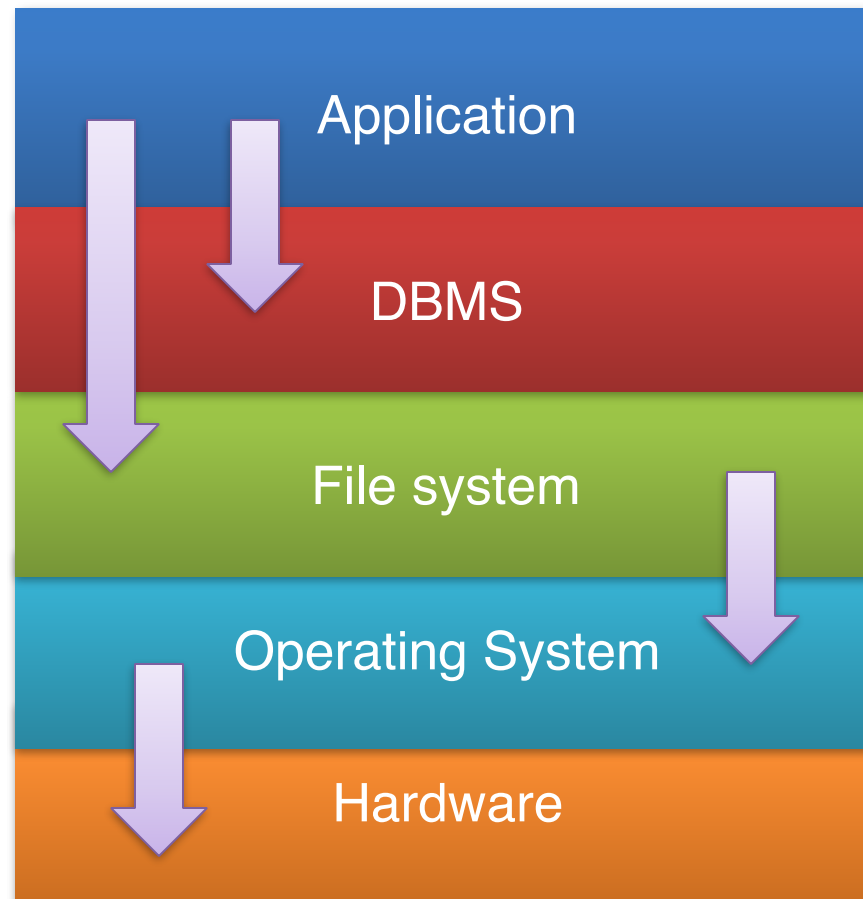


Basic File Processing Operations

- ◆ Opening
- ◆ Closing
- ◆ Reading
- ◆ Writing
- ◆ Seeking
- ◆ Updating
- ◆ Compacting



Where do File System fits in Computer System?



File structures versus DBMS

- ◆ According to Alan Tharp, “file structures is used to process data in physical level, DBMS is used to manage data in a logical level”
- ◆ According to Raghu Ramakrishnan, “DBMS is a piece of software designed to make data maintenance easier, safer, and more reliable”.
- ◆ Thus, file structure is a pre-requisite to DBMSs.



Physical Files and Logical Files

- ◆ **physical file:** a collection of bytes stored on a disk or tape
- ◆ **logical file:** an interface that allows the application programs to access the physical file on the SS
- ◆ The **operating system** is responsible for associating a logical file in a program to a physical file in a SS. Writing to or reading from a file in a program is done through the operating system.



Characteristics of storage

- ◆ Volatility
- ◆ Mutability
- ◆ Accessibility
- ◆ Addressability
- ◆ Performance
- ◆ Capacity



Date	Date	Topics
1	Feb, 22-23	Welcome, DIKW, C intro
2	Mar, 01-02	XML, DTD, XSD
3	Mar, 08-09	XSL and XQuery
4	Mar, 15-16	XML Applications
5	Mar, 22-23	Storage devices & Basic File Ops.
6	Mar, 29-20	Midterm Exam
7	Apr, 05-06	Analytical modelling & Simple Indexing
8	Apr, 12-13	B+tree
9	Apr, 19-20	Multidimensional Indexing
10	Apr, 26-27	External Sorting
11	May, 03-04	ER Model and Normalization
12	May, 10-11	Midterm Exam
13	May, 17-18	DBMS and SQL
14	May, 24-25	Structured Query Language

