City University of Hong Kong

Information on a Course
offered by Department of Computer Science
with effect from Semester A in 2012 / 2013

Part I

Course Title: Data Warehousing and Data Mining

Course Code: CS5483

Course Duration: One Semester

Credit Units: 3

Level: P5

Medium of Instruction: English

Prerequisites: CS3402 Database Systems or equivalent

Precursors: Nil

Equivalent Courses: Nil

Exclusive Courses: Nil

Part II

Course Aims

This course aims to introduce students to a new frontier in database technology, "data warehousing and data mining", by studying their principles, algorithms, implementation methodology, and applications. It will analyze the components of a data warehouse, including data source and transformation tools, metadata management, query reporting and OLAP; provide a comprehensive introduction to data mining, including data selection, cleaning, coding, using different pattern recognition techniques, and reporting; and introduce students to the applications of data warehousing and data mining by using commercial tools for creating business applications.

Course Intended Learning Outcomes (CILOs)
Upon successful completion of this course, students should be able to:

No.	CILOs	Weighting (if applicable)
1.	identify the main characteristics of different data warehousing and data mining techniques through observation of their operations;	
2.	perform a critical assessment of current data warehousing and data mining techniques;	
3.	implement the main algorithms in data warehousing and data mining in a computationally efficient way;	
4.	propose new solutions for data warehousing and data mining problems by improving and combining current techniques.	

Teaching and Learning Activities (TLAs)
(Indicative of likely activities and tasks designed to facilitate students' achievement of the CILOs. Final details will be provided to students in their first week of attendance in this course)

Teaching pattern:
Suggested lecture/tutorial/laboratory mix: 2 hrs. lecture; 1 hr. tutorial.

CILO No.	TLAs	Hours/week (if applicable)
CILO 1,2	**Lecture:** The lecture will focus on the introduction of data warehousing and data mining techniques, and their applications in different problem domains.	
CILO 1,2	**Assignment:** This assignment gives students an opportunity to identify the main characteristics of different techniques in data warehousing and data mining. Students are also required to search literature and perform critical assessment of recent data warehousing and data mining methodologies. This activity supports CILOs 1 and 2.	
CILO 3,4	**Project:** There will be three projects. The first project gives students an opportunity to improve the computational efficiency of existing algorithms in data warehousing and data mining. The second and third projects encourage students to create new designs for data warehousing and data mining systems. This activity supports CILOs 3 and 4.	

Assessment Tasks/Activities

(Indicative of likely activities and tasks designed to assess how well the students achieve the CILOs. Final details will be provided to students in their first week of attendance in this course)

CILO No.	Type of Assessment Tasks/Activities	Weighting (if applicable)	Remarks
CILO 1	**Coursework:** The quality of students' assignments will be used to assess this ILO. **Examination:** Final examination will include questions to assess the capability of students to identify the distinguishing features of different data warehousing and data mining techniques.		
CILO 2	**Coursework:** Students are required to perform critical assessment of different data warehousing and data mining techniques in the assignment. The quality and relevancy of their critiques will be a measure of this ILO. **Examination:** Final examination will include questions to assess the capability of students to perform critical assessment of different data warehousing and data mining methods.		
CILO 3	**Coursework:** In the first project, students will need to improve the computational efficiency of selected data warehousing and data mining algorithms. The quality of the implementation will be a measure of this ILO.		
CILO 4	**Coursework:** In the second and third projects, students will need to identify the limitations of selected data warehousing and data mining systems, and propose new algorithms to address these problems. The creativity and the effectiveness of the proposed designs will be the measures of this ILO. **Examination:** Final examination will include case studies to evaluate students' capability in creating new solutions for challenging problems in data warehousing and data mining.		

Grading of Student Achievement: Refer to Grading of Courses in the Academic Regulations for Taught Postgraduate Degrees.

Examination duration: 2 hours

Percentage of coursework, examination, etc.: 50% CW; 50% Exam

Grading pattern: Standard (A+AA-…F)

For a student to pass the course, at least 30% of the maximum mark for the examination must be obtained.

Part III

Keyword Syllabus
Data extraction, data cleansing, data transformation, metadata, on-line analytical processing (OLAP), star schema, decision trees, neural networks, nearest neighbor and clustering, genetic algorithms, rule induction, data visualization, knowledge discovery in database.

Syllabus
Data Warehousing

XML database, schema translation, schema integration, star schema and data cube, data conversion and intergration, Online Analytical Processing

Data Mining

Association rule, web mining, decision tree, clustering, neural network, genetic algorithm

Recommended Reading
Text(s)
Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques" Third Edition, Morgan Kanfmann Publishers

P.N. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining" Addison Wesley

Joseph Fong, "Information Systems Reengineering and Integration" Second edition Springer Verlag, 2006, ISBN 978-1-84628-382-6

Online Resources