

MODULE DESCRIPTOR

MODULE TITLE	KNOWLEDGE DISCOVERY		
MODULE CODE	CO4762 (L7)	CREDIT VALUE	20 UK CREDITS / 10 ECTS
SCHOOL	SCHOOL OF SCIENCES		

MODULE AIMS

This module considers the concepts and applications of knowledge discovery. It discusses the architecture of data-warehouses and the processes involved in their creation and maintenance. A useful range of important techniques used to data-mine large datasets — to visualise, explore, and find unsuspected patterns, clusters, correlations, connections etc. — are discussed.

The main objectives of the module are to:

- Understand the value of knowledge discovery in solving real-world problems
- Understanding of foundational concepts underlying data mining.
- Evaluate important knowledge discovery techniques
- Apply a wide range of knowledge discovery tools to real-world problems.
- Evaluate the processes involved in the creation and maintenance of data warehouses.

MODULE CONTENT

Indicative syllabus content:

Knowledge Discovery Concepts: Data, Databases, Data Warehouses, Patterns, Technologies, Applications, Research Directions.

Data Preprocessing: Cleaning, Integration, Reduction, Transformation

Data Warehousing and OLAP: Architectures, Modelling, Data Cubes, OLAP, Design and Usage, Implementation

Frequent Patterns, Associations, and Correlations: Basic Concepts, Apriori Algorithm, Generating Association Rules from Frequent Itemsets, Pattern-Growth Approach, Mining Frequent Itemsets Using Vertical Data Format, Mining Closed and Max Patterns, Pattern Evaluation, Advanced Pattern Mining

Classification: Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Model Evaluation and Selection, Support Vector Machines, Neural Networks, Advanced Methods

Cluster Analysis: Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Advanced Methods

Outlier Detection: Outliers and Outlier Analysis, Methods, Statistical Approaches, Statistical Approaches,

Selected Topics in Knowledge Discovery: Personalized exploration, guided analysis, Data Mining and Society

INTENDED LEARNING OUTCOMES

On successful completion of this module a student will be able to:

1. Describe data-mining techniques and assess their applicability
2. Evaluate models/algorithms with respect to their accuracy
3. Critique the results of a data mining exercise
4. Develop hypotheses based on the analysis of the results obtained and test them
5. Make judgements about case studies through the assessment of discovered criteria from data mining

TEACHING METHODS

Lectures deliver factual material, introduce key concepts, direct reading and relate academic aspects to practical considerations.

Tutorial sessions allow students to apply the techniques and reinforce the material delivered in the lecture.

Practical sessions enable students to discuss material and complete online or paper-based exercises.

The module will be assessed by one written coursework. The assignment requires the student to apply specific knowledge discovery techniques on datasets to reveal important findings and summarise, organise and communicate the generated knowledge through a report.

Distance learning

The module tutor will deliver live online lectures through Adobe Connect. During the live lectures the participating students will have the opportunity to engage in discussions, present their views and ask questions. The lecture sessions will be recorded and made available to the students through Blackboard. Students who cannot participate in a live lecture will have the opportunity to answer and reflect on guided questions in the subsequent live lectures or participate asynchronously on discussion boards. The module tutor will provide appropriate feedback to students' comments, as a result of the discussions. Tutor feedback will primarily be provided in an asynchronous manner through Blackboard and emails, but when the need arises, the module tutor will schedule live sessions to provide further feedback. Where appropriate, students will be also provided with relevant further reading, web links and resources for independent study. Speakers from leading organizations will be invited, where possible, to deliver invited talks and enhance the students' experience.

Students will also be provided with bi-weekly self-assessment quizzes, so that they can reflect on their progress.

Students will be provided with access to specialised software/datasets/scripts/programs, through which they will be able to complete the practical components of the module. The students will obtain the practical sheets from Blackboard and they are expected to follow the instructions included in the practical sheets to complete the lab work. If students have difficulties with a particular exercise, they are expected to contact the module tutor or post a question on the discussion forum, where the module tutor and/or their peers can provide feedback. Different means of communication will be utilized by the tutor to offer support to the students based on the reported issue, i.e. email, Skype, Adobe Connect, etc.

ASSESSMENT METHODS

This module is assessed through one Portfolio of coursework and one Examination.