


QFO-AP-FI-MO02	اسم النموذج: Course Syllabus	<p>جامعة فيلادلفيا</p>  <p>Philadelphia University</p>
رقم الاصدار : 1 (Revision)	الجهة المصدرة: كلية تكنولوجيا المعلومات	
التاريخ: 2017/11/05	الجهة المدققة: عمادة التطوير والجودة	
عدد صفحات النموذج:		

Course code: 731423	Course Title: Data Mining and Data Warehousing
Course prerequisite(s) and/or co prerequisite(s): 750361	Course Level: 4
Credit hours: 3	Lecture Time: 11:15 -12:30

Academic Staff Specifics

E-mail Address	Office Hours	Office Number and Location	Rank	Name
immq@yahoo.com	13:00-14:00	305- IT Buliding	Lecturer	Issa Qabaja

Course Description:

This module builds on the introductory module in data warehouse and data mining. It intends to introduce more advanced topics in databases such as data mining and data warehousing.

Course Objectives:

1. Provide the student with an understanding of the concepts of data warehousing and data mining
2. Study the dimensional modeling technique for designing a data warehouse
3. Study data warehouse architectures, OLAP and the project planning aspects in building a data warehouse
4. Explain the knowledge discovery process
5. Describe the data mining tasks and study their well-known techniques
6. Develop an understanding of the role played by knowledge in a diverse range of intelligent systems.
7. Test real data sets using popular data mining tools such as WEKA

Course Components

- 1 Course Overview

- 2 Data Warehouse
- 3 Data preprocessing
- 4 What is data mining?
- 5 Data Mining Techniques
- 6 Cluster analysis
- 7 Distributed & Multimedia Data Bases

Text book:

1. Han, J. and Kamber, M., **Data Mining: Concepts and Techniques**, Morgan Kaufmann, 2012.
2. Tan, P-N, Steinbach, M., Kumar, V. **Introduction to Data Mining**. Addison Wesley, 2006.

IN ADDITION TO THE ABOVE, THE STUDENTS WILL BE PROVIDED WITH HANDOUTS BY THE LECTURER.

Teaching Methods:

Duration: 16 weeks, 48 hours in total

Lectures: 38 hours, 3 hours per week, (including two 1-hour midterm exams)

Laboratory: 4 hours, (0-1 per week, on project assignment)

Seminars: 6 hours (in last 2 weeks)

Learning Outcomes:

- A. To provide a brief introduction to general issues of Data Warehouse and Data Mining.
- B. To provide students with a clear understanding of the different architectures and mining techniques
- C. To introduce students to the role and function of Data Warehouse and Data Mining
- D. To explain the stages and process different data mining techniques.
- E. To learn mining and warehouse techniques through the use of different tools (e.g. ORACLE)

To learn the evaluation techniques of data mining and data warehouse. Assessment Instruments

<u>Allocation of Marks</u>	
Mark	Assessment Instruments
20%	First examination
20%	Second examination
40 %	Final Exam (written unseen exam)
20%	Reports, Assignments, Quizzes, Home works
100%	Total

** MAKE-UP EXAMS WILL BE OFFERED FOR VALID REASONS ONLY WITH CONSENT OF THE DEAN. MAKE-UP EXAMS MAY BE DIFFERENT FROM REGULAR EXAMS IN CONTENT AND FORMAT.*

Practical Submissions

The assignments that have work to be assessed will be given to the students in separate documents including the due date and appropriate reading material.

Documentation and Academic Honesty

Submit your home work covered with a sheet containing your name, number, course title and number, and type and number of the home work (e.g. tutorial, assignment, and project).

Any completed homework must be handed in to my office (room IT ---) by 15:00 on the due date. After the deadline “zero” will be awarded. You must keep a duplicate copy of your work because it may be needed while the original is being marked.

You should hand in with your assignments:

- 3 A printed listing of your test programs (if any).
- 4 A brief report to explain your findings.
- 5 Your solution of questions.

For the research report, you are required to write a report similar to a research paper. It should include:

- **Abstract:** It describes the main synopsis of your paper.
- **Introduction:** It provides background information necessary to understand the research and getting readers interested in your subject. The introduction is where you put your problem in context and is likely where the bulk of your sources will appear.
- **Methods (Algorithms and Implementation):** Describe your methods here. Summarize the algorithms generally, highlight features relevant to your project, and refer readers to your references for further details.
- **Results and Discussion (Benchmarking and Analysis):** This section is the most important part of your paper. It is here that you demonstrate the work you have accomplished on this project and explain its significance. The quality of your analysis will impact your final grade more than any other component on the paper. You should therefore plan to spend the bulk of your project time not just gathering data, but determining what it ultimately means and deciding how best to showcase these findings.
- **Conclusion:** The conclusion should give your reader the points to “take home” from your paper. It should state clearly what your results demonstrate about the problem you were tackling in the paper. It should also generalize your findings, putting them into a useful context that can be built upon. All generalizations should be supported by your data, however; the discussion should prove these points, so that when the reader gets to the conclusion, the statements are logical and seem self-evident.
- **Bibliography:** Refer to any reference that you used in your assignment. Citations in the body of the paper should refer to a bibliography at the end of the paper.

Protection by Copyright

1. Coursework, laboratory exercises, reports, and essays submitted for assessment must be your own work, unless in the case of group projects a joint effort is expected and is indicated as such.
2. Use of quotations or data from the work of others is entirely acceptable, and is often very valuable provided that the source of the quotation or data is given. Failure to provide a source or put quotation marks around material that is taken from elsewhere gives the appearance that the comments are ostensibly your own. When quoting word-for-word from the work of another person quotation marks or indenting (setting the quotation in from the margin) must be used and the source of the quoted material must be acknowledged.
3. Sources of quotations used should be listed in full in a bibliography at the end of your piece of work.

Avoiding Plagiarism.

1 Unacknowledged direct copying from the work of another person, or the close paraphrasing of somebody else's work, is called plagiarism and is a serious offence, equated with cheating in examinations. This applies to copying both from other students' work and from published sources such as books, reports or journal articles.

2 Paraphrasing, when the original statement is still identifiable and has no acknowledgement, is plagiarism. A close paraphrase of another person's work must have an acknowledgement to the source. It is not acceptable for you to put together unacknowledged passages from the same or from different sources linking these together with a few words or sentences of your own and changing a few words from the original text: this is regarded as over-dependence on other sources, which is a form of plagiarism.

3 Direct quotations from an earlier piece of your own work, if not attributed, suggest that your work is original, when in fact it is not. The direct copying of one's own writings qualifies as plagiarism if the fact that the work has been or is to be presented elsewhere is not acknowledged.

4 Plagiarism is a serious offence and will always result in imposition of a penalty. In deciding upon the penalty the Department will take into account factors such as the year of study, the extent and proportion of the work that has been plagiarized, and the apparent intent of the student. The penalties that can be imposed range from a minimum of a zero mark for the work (without allowing resubmission) through caution to disciplinary measures (such as suspension or expulsion).

Course Academic Calendar

Week	Basic and support material to be covered	Homework/reports and their due dates
(1)	Course Overview Course Introduction Knowledge discovery process Why data warehouse & data mining	
(2)	Data Warehouse Why data warehouse? OLTP and OLAP Data Cube Data Warehouse modeling Warehouse views Data Warehouse Architectures Data Warehouse implementation	
(3) & (4)	Data preprocessing Why preprocess the data? Data cleaning Data integration and transformation Data reduction Dimensionality reduction Data compression Feature extraction Discretization and concept hierarchy generation	
(5)	Applications on Data Warehouse Case Study Assignment 1	Assignment 1
(6)	What is data mining Motivation and challenges of data mining Data mining tasks Types of Data Data set types Data mining applications	
(7)	Data quality Data preprocessing: Aggression, sampling, dimensionality reduction, feature selection, feature creation, discretization, transformation Measuring the similarity and dissimilarity between: Simple attributes, data objects	
(8)	Proximity measures Issues in proximity calculation	
(9) & (10)	Data Mining Techniques Mining association rules Association rule mining Apriori algorithm Frequent Pattern Growth algorithm Rule based Classification What is classification	

	Decision trees: ID3, C4.5 Rule induction: RIPPER algorithm	
(11)	Data Mining Techniques Rule based Classification Associative classification (CBA, MMAC) Rule Pruning : REP, database coverage	
(12)	Data Mining Techniques Other classification approaches Regression Neural networks Genetic algorithms Cluster analysis Partitioning methods (K-means) Hierarchical methods	Assignment 2
(13)	Distributed Data Base	
(14)	Multimedia Data Base Parallel Data Base	
(15)	Case Study: Text Categorization Tutorial: Using associative classification for text categorization	
(16)	review questions and Final Exam	

Expected workload:

On average students need to spend 2 hours of study and preparation for each 50-minute lecture/tutorial.

Attendance Policy:

Absence from lectures and/or tutorials shall not exceed 15%. Students who exceed the 15% limit without a medical or emergency excuse acceptable to and approved by the Dean of the relevant college/faculty shall not be allowed to take the final examination and shall receive a mark of zero for the course. If the excuse is approved by the Dean, the student shall be considered to have withdrawn from the course.

Module References

1. Connolly, T., and Begg, C. Data Base Systems: A Practical Approach to Design, Implementation, and Management, Addison Wesley, fourth edition, 1999.
2. Witten, I., and Frank, E. mining: practical machine learning tools and techniques with Java implementations. San Francisco: Morgan Kaufmann, 2001.