SUBJECT DESCRIPTION FORM

Subject Title: Data Mining and Data Warehousing Applications

Subject Code: COMP5121

Credit Value: 3

Pre-requisite: (Subject title and code no, if any)

Nil

Recommended background knowledge:

Nil

Mutual Exclusions:

Data Mining and Applications (COMP518),
Data Mining and Data Warehouse (COMP578),
Data Mining and Data Warehousing for Business (COMP5003)

Learning Approach:

42 hours of class activities including - lecture, tutorial, lab, workshop seminar where applicable

There will be a mix of lectures, discussions and case study analysis. Recent articles in the area of data mining and data warehousing strategies will also be reviewed and discussed in lectures. Students will be given the flexibility to tailor the study material to their organization environment and pursue relevant case studies.

Assessment:

Continuous Assessment	60%
Test, and Examination	40%

Objectives:

- make more effective use of data stored in data bases
- create a clean, consistent repository of data within a data warehouse
- utilize various levels and types of summarization of data to support management decision making
- discover patterns and knowledge that is embedded in the data using different data mining techniques

The Department reserves the right to update the syllabus contents. Please note that the learning approach for the same subject could vary slightly due to different delivery modes.

Learning Outcomes:

After completing this subject, students should be able to:

1. understand the need for data warehouse;
2. identify components in typical data warehouse architecture;
3. design a data warehouse in support of business problem solving;
4. understand typical knowledge discovery process and the different algorithms available by popular commercial data mining software; and
5. obtain hands-on experience with some popular data mining software.

Keyword syllabus:

Students will be presented with documentation (lecture notes, tutorial materials, practical exercises and reference papers) of a range of up-to-date methods, techniques and applications in data mining and data warehouse. More specifically, this subject covers the following topics:

- Introduction to data warehousing and data mining; possible application areas in business and finance; definitions and terminologies; types of data mining problems.
- Data warehouse and data warehousing; data warehouse and the industry; definitions; operational databases vs. data warehouses.
- Data warehouse architecture and design; two-tier and three-tier architecture; star schema and snowflake schema; data characteristics; static and dynamic data; meta-data; data marts.
- Data replication, data capturing and indexing, data transformation and cleansing; replicated data and derived data; Online Analytical Processing (OLAP); multidimensional databases; data cube
- Data Mining and knowledge discovery, the data mining lifecycle; pre-processing; data transformation; types of problems and applications.
- Mining of Association Rules; the Apriori algorithm; binary, quantitative and generalized association rules; interestingness measures.
- Classification; decision tree based algorithms; Bayesian approach; statistical approaches, nearest neighbor approach; neural network based approach; Genetic Algorithms based technique; evaluation of classification model.
- Clustering; k-means algorithm; Hierarchical algorithm; Condorset; neural network and Genetic Algorithms based approach; evaluation of effectiveness.
- Sequential data mining; time dependent data and temporal data; time series analysis; sub-sequence matching; classification and clustering of temporal data; prediction.
- Computational intelligence techniques; fuzzy logic, genetic algorithms and neural networks for data mining.

Indicative reading list and references:

Books

Han, J., and Kamber, M., 2006, *Data Mining: Concepts and Techniques*, 2nd Ed., Morgan Kaufmann, San Francisco, CA.

Tan, P.N., Steinbach, M., Kumar V., *Introduction to Data Mining*, Addison Wesley, 2006.

Dunham, M.H., *Data Mining: Introductory and Advanced Topics*. Prentice Hall, 2003.

Chakrabarti, S., *Mining the Web*. Morgan Kaufmann Publishers, 2003.

Chan, K.C.C., 2002, *Course Notes on Data Mining & Data Warehousing*, Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

Inmon, W.H., 1996, *Building the Data Warehouse*, 2nd Edition, J. Wiley & Sons, New York, NY.

Mattison, R., 1997, *Data Warehousing and Data Mining for Telecommunications*, Artech House, Boston.

Wtehorn, M., 1999, *Business Intelligence: the IBM Solution: Datawarehousing and OLAP*, Springer, London.
Pyle, D., *Data Preparation for Data Mining*. Morgan Kaufmann Publishers, 1999.
Rud, 2001, *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management*, J. Wiley, New York, NY.
Groth, R., 1998, *Data Mining: Building Competitive Advantage*, Prentice Hall, Upper Saddle River, NJ.
Berry, M.J.A., 2000, *Mastering Data Mining: the Art and Science of Customer Relationship Management*, Wiley, New York NY.
Kovalerchuk, B., 2000, *Data Mining in Finance: Advances in Relational and Hybrid Methods*, Kluwer Academic, Boston.
Berry, M.J.A., 1997, *Data Mining Techniques for Marketing, Sales and Customer Support*, Wiley, New York NY.