

C. 1997
(p) ✓

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

1. DEPARTAMENTO: Computación
 2. CUATRIMESTRE: Segundo de 1997.
 3. ASIGNATURA: DATA WAREHOUSING
 4. CARRERA: Licenciatura en Ciencias de la Computación
 5. CARACTER DE LA MATERIA: Optativa
 6. NUMERO DE CODIGO DE CARRERA: 18
 7. NUMERO DE CODIGO DE MATERIA: C039
 8. PUNTAJE: 3 puntos (planes 87 y 93)
 9. PLAN DE ESTUDIOS AÑO: 1987 y 1993.
 10. DURACION DE LA MATERIA: Cuatrimestral
 11. HORAS DE CLASE SEMANAL:
 - a) TEORICAS 3 HS.
 - b) LABORATORIO
 - c) PROBLEMAS 3 HS.
 - d) SEMINARIOS
 12. CARGA HORARIA TOTAL: 6 HORAS
 13. ASIGNATURAS CORRELATIVAS: Base de Datos (planes 87 y 93)
 14. FORMA DE EVALUACION: Examen Final
 15. PROGRAMA Y BIBLIOGRAFIA: Adjuntas a esta hoja
- FECHA: 1/11/97

JA
Lic. Juan Ale
Firma y Aclaración
del Profesor Titular

Firma del Director
y Sello Aclaratorio

Lic. IRENE LOISEAU
DIRECTORA
DEPTO. DE COMPUTACION
F. C. E. y N. UBA



Departamento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

DATA WAREHOUSING

2º Cuatrimestre de 1997

Objetivo: proveer los conceptos y elementos necesarios para el diseño e implementación de un data warehouse, como así también técnicas para su explotación. Por otra parte, se presentarán tópicos que son objeto de investigación actual.

Contenido

1. Conceptos básicos de Data Warehousing. Diferencias entre OLTP y OLAP. Arquitectura e infraestructura técnica.
2. Metodología para el diseño de DSS. Diseño de bases de datos multidimensionales: Star Schema, Snowflake Schema. Diferencias con metodologías tradicionales de diseño de bases de datos. Optimización del diseño. Aplicaciones. Estudio de casos.
3. Implementación del Data Warehouse: procesos de extracción, cleaning, formateo y carga de datos. Técnicas de refreshment, sumarización y agregación. Definición y gestión del Metadata.
4. Enfoques MD-OLAP y R-OLAP. Comparación de los enfoques. Herramientas de acceso para el usuario final. Productos comerciales.
5. Tópicos actuales de investigación: materialización de views, optimización de consultas complejas con agregados, visión del Data Warehouse como un Data Cube.
6. Explotación del Data Warehouse: Diversos enfoques y técnicas. Análisis multidimensional. Descubrimiento del Conocimiento (KDD).
7. Técnicas de Data Mining. Descubrimiento de diversos tipos de conocimiento: Análisis de algoritmos para reglas de asociación, clasificación, clustering, etc. El problema de "lo interesante". Aspectos de eficiencia y escalabilidad. Problemas no resueltos.

Metodología de la enseñanza: clases teórico-prácticas, con fuerte interacción entre estudiantes y docentes. Se desarrollará un trabajo de laboratorio en el que se usará un producto comercial.

Carga horaria semanal: 2 reuniones de 3 horas cada una.

Prerrequisitos: T.P. de Bases de Datos aprobado.

Aprobación: aprobación del trabajo de laboratorio, un coloquio y examen final.

Profesor: Lic. Juan M. Ale

Docentes Auxiliares: Lic. Alejandro Vaisman
Lic. Alejandro Eidelsztejn

3 puntos

Lic. IRENE LOISEAU
DIRECTORA
DEPTO. DE COMPUTACION
F. C. E. y N. UBA

Bibliografía sobre Data Warehouse y Data Mining:

Libros

- 1 - Adriaans, P.-Zantinge, D.: Data Mining. Addison-Wesley. 1996.
- 2 - Barquin: Planning & Designing the Data Warehouse. Wiley. 1997.
- 3 - Bigus: Data Mining with Neural Networks. McGraw-Hill. 1996.
- 4 - Devlin: Data Warehouse: from Architecture to Implementation. Addison-Wesley. 1997.
- 5 - Fayyad: Advances in Knowledge Discovery and Data Mining. The MIT Press. 1996.
- 6 - Gill: The Official Guide to Data Warehousing. QUE. 1996.
- 7 - Inmon: Building the Data Warehouse. Wiley. 2da. Ed. 1996.
- 8 - Inmon: Managing the Data Warehouse. Wiley. 1997.
- 9 - Kimball: The Data Warehouse Toolkit. Wiley. 1996.
- 10 - Mattison: Data Warehousing. Wiley. 1996.
- 11 - Poe: Building a Data Warehouse for Decision Support. Prentice Hall. 1996.

Artículos sobre Data Mining

- [AHO96]. Apte, C.-Hong, S.J.: Predicting Equity Returns from Securities Data with Minimal Rule Generation. In Advances in KD and DM. The MIT Press. 1996.
- [AIS93]. Agrawal, R.-Imielinski, T.-Swami, A.: Mining Association Rules Between Sets of Items in Large Databases. Proc. ACM SIGMOD. 1993.
- [AMS96]. Agrawal, R.-Mannila, H.-Srikant, R.-Toivonen, H.-Verkano, I.: Fast Discovery of Association Rules. In Advances in KD and DM. The MIT Press. 1996.
- [ASh96]. Agrawal, R.-Shafer, J.C.: Parallel Mining of Association Rules. IEEE TOKDE Vol.8 n°6. Dec. 1996.
- [BAn96]. Brachman, R.-Anand, T.: The Process of Knowledge Discovery in Databases: A Human-Centered Approach. In Advances in KD and DM. The MIT Press, 1996.
- [BCI96]. Berndt, D. Clifford, J.: Finding Patterns in Time Series: A Dynamic Programming Approach. In Advances in KD and DM. The MIT Press. 1996.
- [BKK96]. Brachman, R.-Khabaza, T.-Kloesgen, W.-Piatetsky-Shapiro, G.-Simoudis, E.: Mining Business Databases. ACM Communications Vol.39 N°11. Nov. 1996.
- [Bun96]. Buntine, Wray: Graphical Models for Discovering Knowledge. In Advances in KD and DM. The MIT Press 1996.

- [CHY96]. Chen, M.S.-Han, J.-YU, P.S.: Data Mining: An Overview from a Database Perspective. IEEE TOKDE Vol. 8 N° 6. Dec. 1996.
- [CNF96]. Cheung, D.W.-Ng, V.T.-Fu, A.W.-Fu, Y.: Efficient Mining of Association Rules in Distributed Databases. IEEE TOKDE Vol. 8 N°6. Dec. 1996.
- [CSt96]. Cheeseman, P.-Stutz, J.: Bayesian Classification (AutoClass): Theory and Results. In Advances in KD and DM. The MIT Press. 1996.
- [Dze96]. Dzeroski, Saso: Inductive Logic Programming and Knowledge Discovery in Databases. In Advances in KD and DM. The MIT Press. 1996
- [EPr96]. Elder, J.-Pregibon, D.: A Statistical Perspective on Knowledge Discovery in Databases. In Advances in KD and DM. The MIT Press. 1996.
- [Etz96]. Etzioni, O.: The World-Wide Web: Quagmire or Gold Mine?. ACM Communications. ACM Communications Vol.39 N°11. Nov. 1996.
- [FDW96]. Fayyad, U.M.-Djorgovski, S.G.-Weir, N.: Automating the Analysis and Cataloging of sky Surveys. In Advances of KD and DM. The MIT Press. 1996.
- [FHS96]. Fayyad, U.-Haussler, D.-Stolorz, P.: Mining Scientific Data. ACM Communications Vol.39 N°11. Nov. 1996.
- [FLi95]. Faloutsos, C.-Lin, K.: Fast-Map: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. Proc. ACM SIGMOD. 1995.
- [FMM96]. Fukuda, T.-Morimoto, Y.-Morishita, S.-Tokuyama, T.: Data Mining Using Two-Dimensional Optimized Association Rules: Schema, Algorithms, and Visualization. Proc. ACM SIGMOD. 1996.
- [FPS96a]. Fayyad, U.-Piatetsky-Shapiro, G.-Smyth, P.: From Data Mining to Knowledge Discovery. In Advances in KD and Data Mining. The MIT Press. 1996.
- [FPS96b]. Fayyad, U.-Piatetsky-Shapiro, G.-Smyth, P.: The KDD Process for Extracting Useful Knowledge From Volumes of Data. ACM Communications Vol.39 N°11. Nov. 1996.
- [Gai96]. Gaines, Brian: Transforming Rules and Trees into Comprehensible Knowledge Structures. In Advances in KD and DM. The MIT Press. 1996.
- [GGT94]. Gravano, L.-Garcia-Molina, H.-Tomasic, A.: The Effectiveness of GLOSS for the Text Database Discovery Problem. Proc. ACM SIGMOD. 1994.
- [GMP96]. Glymour, C.-Madigan, D.-Pregibon, D.-Smyth, P.: Statistical Inference and Data Mining. ACM Communications Vol.39 N°11. Nov. 1996.
- [GMV96]. Guyon, I.-Matic, N.-Vapnik, V.: Discovering Informative Patterns and Data Cleaning. In Advances in KD and DM. The MIT Press. 1996.
- [GTN96]. Goh, C.L.-Tsukamoto, M.-Nishio, S.: Knowledge Discovery in Deductive Databases with Large Deduction Results: The First Step. IEEE TOKDE Vol.8 N°6. Dec. 1996.
- [Hec96]. Heckerman, D.: Bayesian Networks for Knowledge Discovery. In Advances in KD and DM. The MIT Press. 1996.

- Genesereth, M.R., Nilsson, N.J.: Logical Foundations of Artificial Intelligence, Los Altos, Morgan Kaufmann, 1987.
- Goldberg, D.E.: Genetic Algorithms in search, optimization and machine learning, Reading, Addison-Wesley, 1989.
- Holland, J.H., Holyoak, K.J., Nisbett, R.E., Thagard, P.R.: Induction - Processes of Inference, Learning and Discovery, Cambridge, MIT Press, 1989.
- Kahneman, D., Slovic, P. y Tversky, A.: Judgment under Uncertainty: Heuristics and Biases. New York: Cambridge University Press. 1982.
- Michalski, R., Carbonell, J. y Mitchell, T. (eds.) Machine Learning: An Artificial Intelligent Approach, Palo Alto, Ca., Tioga Publishing Company, 1983.
- Nilsson, N.J.: Principles of Artificial Intelligence, Palo Alto, Tioga Publishing, 1980.
- Pearl, J.: Heuristics - Intelligent Search Strategies for Computer Problem Solving, Reading, Addison-Wesley, 1984.
- Pearl, Judea: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, San Francisco, Morgan Kaufmann, 1988.
- Rich, E., Knight, K.: Artificial Intelligence, New York, Mc-Graw-Hill, 1991.
- Shafer, G.: A mathematical theory of evidence, New Jersey, Princeton University Press, 1976.
- Winston, P.H.: Inteligencia Artificial, Wilmington, Addison-Wesley, 1994.



Lic. IRENE LOISEAU
DIRECTORA
DEPTO. DE COMPUTACION
F. C. E. y N. UBA