

**PROGRAMMA DEL CORSO DI
BASI DI DATI E BASI DI CONOSCENZA**

A.A. 2002/2003

Prof. Donato Malerba

Obiettivi. Il corso presenta i concetti e le metodologie evolute per lo sviluppo di sistemi software orientati alla gestione sia dei dati e sia della conoscenza, nei quali le problematiche fondamentali riguardano l'organizzazione, la manipolazione, l'accesso a dati e conoscenza, come pure l'estrazione di conoscenza dai dati.

Il data base di riferimento è Oracle 8.

Prerequisiti: algoritmi e strutture dati, basi di dati e sistemi informativi, ingegneria della conoscenza e sistemi esperti.

Modalità d'esame: prova orale e discussione di un progetto svolto su argomento concordato con il docente (che va contattato con congruo anticipo rispetto agli appelli di esame).

Programma del corso a.a. 2002-2003

1. Le basi di dati attive.

Basi di dati e sistemi a produzione. Comportamento dei trigger in un sistema relazionale. Definizione e uso dei trigger in Oracle. Caratteristiche evolute delle regole attive. Proprietà delle regole attive (terminazione, confluenza e determinismo delle osservazioni). Problemi di progettazione e realizzazione di basi di dati attive. La metodologia IDEA. Applicazioni delle basi di dati attive.

2. Le basi di dati deduttive

Regole attive e deduttive. Regole deduttive e viste logiche. SQL3 e ricorsione. Database deduttivo relazionale. Datalog. Teoria dei modelli. Modello Minimo di Herbrand. Datalog e algebra relazionale. La strategia di ricerca. Valutazione bottom-up. Valutazione top-down. Magic set. La negazione in Datalog. Datalog⁺ stratificato. Estensioni del Datalog per trattare oggetti complessi. Proprietà di database Datalog. Sistemi di basi deduttive e sistemi di programmazione logica. Approcci architetturali alle basi di dati deduttive.

3. Le basi di dati relazionali e a oggetti.

Modelli dei dati non in prima forma normale. Il modello relazionale a oggetti. SQL-3: tuple e oggetti, gerarchie di tipo, tipi astratti, interrogazioni con flattening e nesting. Il manifesto delle basi di dati della terza generazione. Un DBMS relazionale a oggetti: Illustra. L'implementazione a oggetti di Oracle 8: tipi di dati astratti, collezioni, oggetti riga, viste oggetto. Basi di dati a oggetti: tipi, classi, metodi, gerarchie di generalizzazione, persistenza, ridefinizioni dei metodi, ridefinizione con raffinamenti di tipo. Il manifesto delle basi di dati a oggetti. Standard ODMG per le basi di dati a oggetti: ODL e OQL. Un OODBMS: ObjectStore.

4. Le basi di dati multimediali e spaziali.

I sistemi informatici multimediali. I dati multimediali. Gestione dei dati multimediali. Richiamo e precisione. I sistemi di recupero di informazioni testuali. Indicizzazione manuale e automatica. Creazione degli indici. Funzionalità e progetto di un IRS per dati testuali: il sistema WebClass. I sistemi di recupero di immagini: classificazione e recupero. Funzionalità e progetto di un IRS per immagini di documenti: il sistema WISDOM++. Rappresentazione di oggetti spaziali. Estensione dello SQL per dati spaziali.

5. Basi di dati e World Wide Web.

Richiami su Internet e World Wide Web: nodi, indirizzi e protocolli, il World Wide Web, ancore e URL, HTML, il protocollo http. I sistemi informativi su Web: pubblicazione e consultazione sul web – comunicazione unilaterale, transazioni sul web – comunicazione bidirezionale, commercio elettronico e

altre applicazioni. L'accesso alle basi di dati: accesso a database mediante common gateway interface (CGI), l'interazione field-oriented, vantaggi e svantaggi dei DBMS Web, il server universale. Il Web e la basi di dati: differenze e convergenze.

6. Architetture e modelli per l'analisi dei dati.

Dati operazionali e dati decisionali. Tecnologie di Business Intelligence. Sistemi di supporto alle decisioni (DSS), Executive Information Systems (EIS) e Management Information Systems (MIS). Caratteristiche di un data warehouse. Architettura di un data warehouse. Il modello multidimensionale. Schema del data warehouse: a stella, a fiocco di neve, a costellazione. OLAP e operazioni per l'analisi dei dati: drill down e roll up. ROLAP e MOLAP. Uno studio di caso: il modello relazionale di un DW per il settore agro/alimentare.

7. Scoperta di conoscenza nelle basi di dati.

La scoperta di conoscenza nelle basi di dati: definizione e problemi. Il processo della scoperta di conoscenza nelle basi di dati: la selezione, il preprocessing, la trasformazione, il data mining, l'interpretazione e la valutazione dei risultati. Data mining: obiettivi, task, rappresentazione del modello, la valutazione del modello, i metodi di ricerca. Algoritmi di data mining per compiti descrittivi e predittivi: mining di regole di associazione, analisi di cluster, classificazione e regressione. La costruzione di alberi di decisione. Valutazione dei risultati ottenuti. Linguaggi di interrogazione per il data mining.

I seguenti argomenti sono inoltre stati trattati nell'attività seminariale del Dr. Willi Klösgen:

- Introduzione all'apprendimento automatico
- Metodi elementari di data mining
- Metodi avanzati di data mining
- Estrazione di sottogruppi
- Analisi di dati spaziali
- Analisi di dati su web

8. Il DBMS Oracle 8.

Gli strumenti di Oracle 8. Creazione di database in Oracle. La progettazione concettuale mediante Oracle Designer. Generazione del modello logico mediante Oracle Design Transformer. Generazione di form, di interrogazioni e di trigger mediante Oracle Developer. Creazione di tipi oggetto e interrogazioni. Le estensioni di Oracle per trattare i tipi spaziali. Il modello dei dati spaziale. Le operazioni spaziali. Indicizzazione spaziale e elaborazione di interrogazioni.

Principali testi e articoli di riferimento

P. Atzeni, S. Ceri, S. Paraboschi & R. Torlone

Basi di dati: concetti, linguaggi e architetture (2a edizione)

McGraw-Hill Libri Italia, 1999.

Capitoli: 11.3-11.4, 12, 13, 14, Appendici A e D.

A. Albano, G. Ghelli, R. Orsini

Basi di dati relazionali e a oggetti

Zanichelli, 1997

Capitoli: 10, 11.1, 11.2

J. Wisdom & S. Ceri (Eds.),

Active Database Systems,

Morgan Kaufmann, 1996

S. Ceri, G. Gottlob & L. Tanca

What you always wanted to know about Datalog (and never dared to ask)

IEEE Trans. on Knowledge and Data Engineering, n.1, 1989

E. Bertino, B. Catania & GP. Zarri,

Intelligent Database Systems, Addison-Wesley, 2001,

Capitoli: 3.225, 3.23, 3.24, 3.25

U. Fayyad, G. Piatetsky-Shapiro, P. Smyth. From data mining to knowledge discovery: an overview.

In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining*,

AAAI/MIT Press, pp. 1-35, 1996.

W.J. Frawley, G. Piatetsky-Shapito, C.J. Matheus. Knowledge discovery in databases.

In W.J. Frawley, G. Piatetsky-Shapito, C.J. Matheus, *Knowledge discovery in databases*,

AAAI/MIT Press, pp. 1-27, 1991.

T. Mitchell

Machine Learning

McGraw Hill, 1997.

Capitoli 1 e 3.

J. Han, M. Kamber

Data mining: concepts and techniques

Morgan Kaufmann, 2001

Capitoli: 4,5,6,7,8,9

Copia delle trasparenze proiettate durante le lezioni relative alle parti 1-6, 8 sono disponibili sul sito:

<http://www.di.uniba.it/~malerba/courses/bdbc.htm>

Copia delle trasparenze relative alla parte 7 è disponibile presso la Biblioteca del Dipartimento di Informatica.