

PROGRAMMA DEL CORSO DI
BASI DI DATI AVANZATE
Corso di Laurea Triennale in Informatica – Università degli Studi di Bari
A.A. 2005/2006
Prof. Donato Malerba

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

Il DBMS di riferimento è Oracle 10g.

Prerequisiti: algoritmi e strutture dati, basi di dati.

Modalità d'esame: prova scritta e discussione orale del progetto completo svolto in laboratorio su DBMS Oracle 10g (le specifiche sono disponibili sul sito del corso).

Programma preliminare del corso

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 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, ereditarietà di tipo, identificazione del tipo al run-time.

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 ritrovamento di informazioni testuali: classificazione e recupero di documento. I sistemi di ritrovamento di immagini: classificazione e recupero di immagini. Rappresentazione di oggetti spaziali. Estensione dello SQL per dati spaziali.

5. Basi di dati per il supporto alle decisioni.

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.

6. 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. Valutazione dei risultati ottenuti. Induzione di alberi di decisione.

7. Il DBMS Oracle 10g.

Gli strumenti di Oracle 10g. 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 dati testuali (Oracle Text) e spaziali (Spatial Oracle).

Principali testi e articoli di riferimento

Parti 1,3,4,5

P. Atzeni, S. Ceri, P. Fraternali, S. Paraboschi & R. Torlone
Basi di dati: Architetture e linee di evoluzione
McGraw-Hill Libri Italia, 2003.
Capitoli: 5.3-5.4, 7, 8

Per integrazioni:

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

Parte 2

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

Per integrazioni:

E. Bertino, B. Catania & GP. Zarri,
Intelligent Database Systems, Addison-Wesley, 2001,
Capitoli: 3.225, 3.23, 3.24, 3.25

Parte 6

Richard J. Roiger, Michael W. Geatz.
Introduzione al Data Mining.
McGraw-Hill, 2003
Capitoli 1, 2, 3.1-3.2, 4, 5

Copia delle trasparenze proiettate durante le lezioni e durante le esercitazioni in laboratorio sono disponibili sul sito:

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