Gestione e condivisione di basi di dati e conoscenza

Modulo 2

Federica Cena
cena@di.unito.it

Orario del corso

- lunedì e mercoledì 10.00-13.00
- Da lunedì 6 novembre 2017 a mercoledì 20 dicembre **TRANNE LUNEDI 27 novembre**

Esame

- Presentazione di un progetto: discussione
- Domande di teoria su tutto il programma

- gruppi di due persone massimo
- dividervi già adesso in gruppi in modo da seguire piu' agilmente il corso (progetto costruito via via durante il corso)

Testi

- Alessandro Rezzani, Business Intelligence, Apogeo
→ capitoli 1-2-4-5-6-8-9-11-12 (no 3,7,10)

- C. Berry, B. Linoff: Data mining, Apogeo 2003
→ capitolo 1 e 5

<http://www.di.unito.it/~cena/teaching.html>

Ricevimento

- ▣ Su appuntamento via email
- ▣ cena@di.unito.it

Business Intelligence

metodi e strumenti per convertire

- ▣ dati in informazioni
- ▣ informazioni in conoscenza
- ▣ conoscenza in piani di sviluppo

(D. Loshin 2003)

Gestione e condivisione di basi di dati e di conoscenza

L'obiettivo del corso è quello di fornire una panoramica delle metodologie e applicazioni informatiche avanzate maggiormente utilizzate

- per strutturare dati
 - Data base (Microsoft access, Mysql)
 - Data warehouse (Microsoft access , Mysql)
- analizzare dati
 - OLAP (Online Analytical Processing) (SPSS, Excell, SQL Server analitics, etc)
 - DATA MINING (Weka)

Gestione e condivisione di basi di dati e di conoscenza

Conoscere e saper applicare le tecniche di **modellizzazione semantica** della conoscenza

Conoscere gli **open data** ed essere in grado di pubblicare dati come open, oltre che usare le banche dati open esistenti

Programma del corso

- Prima parte
 - Data warehousing e analisi OLAP
 - Data mining

Programma del corso

- Seconda parte
 - Semantic web,
 - linked open data
 - (Big data) NO

Schematizzando il corso

Strutturazione dei dati		Analisi dei dati	
data base (1 modulo)	data warehouse	OLAP	data mining

Schematizzando il corso

Strutturazione dati → conoscenza	
Linked (open) data (rdf)	Ontologie (Owl)

12

Template progetto d'esame

Progetto

- ▣ argomento di studio
- ▣ obiettivi di analisi

1 parte

Data warehouse:

- schema e-r
- implementazione in mysql
- popolamento con dati (inseriti o importati)
- esportazione come cvs
- **cubi olap:** con excel (tabelle pivot)

2 parte

Analisi di data mining

A partire dai dati del proprio datawarehouse esportati come cvs

- **alberi:** con weka
- **cluster:** con weka

3 parte

▣ **LINKED OPEN DATA e Semantic Web**

- ▣ Annotazione di una (porzione) pagina sito web con RDFa/microdati/Json-LD (a scelta) in base a un vocabolario/ontologia pubblico (ad esempio schema.org, dublincore.org, xmlns.com/foaf/spec/)