

Complementi di Basi di dati – Esame del 20 aprile 2007
Tempo a disposizione: 2 ore (libri chiusi)

Domanda 1 (10%)

Scrivere

- (a) un documento XML valido e
- (b) un documento XML ben-formato ma non valido

rispetto a seguente DTD:

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT autorimessa (auto+)>
<!ELEMENT auto (marca,modello,anno,proprietario)>
<!ELEMENT marca (#PCDATA)>
<!ELEMENT modello (#PCDATA)>
<!ELEMENT anno (#PCDATA)>
<!ELEMENT proprietario EMPTY>
  <!ATTLIST proprietario nome CDATA #IMPLIED>
  <!ATTLIST proprietario cognome CDATA #IMPLIED>
  <!ATTLIST auto targa ID #REQUIRED>
```

Possibile Soluzione

(a)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE autorimessa SYSTEM "auto.dtd">
<autorimessa>
  <auto targa="AA234BB">
    <marca>Peugeot</marca>
    <modello>206</modello>
    <anno>2005</anno>
    <proprietario nome="Mario" cognome="Rossi"/>
  </auto>
  <auto targa="CC432GG">
    <marca>Lancia</marca>
    <modello>Y</modello>
    <anno>2005</anno>
    <proprietario nome="Anna"/>
  </auto>
  <auto targa="YY786DD">
    <marca>Fiat</marca>
    <modello>Punto</modello>
    <anno>2005</anno>
    <proprietario cognome="Nasi"/>
  </auto>
</autorimessa>
```

(b)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE autorimessa SYSTEM "auto.dtd">
<garage>
  <auto targa="AA234BB">
    <marca>Peugeot</marca>
    <modello>206</modello>
  </auto>
</garage>
```

Domanda 2 (20%)

Scrivere l'XML Schema che consente di validare il seguente documento XML:

```
<?xml version="1.0" encoding="UTF-8"?>
<risultati_esame>
  <esame>
    <data>
      <giorno>27</giorno>
      <mese>aprile</mese>
      <anno>2006</anno>
    </data>
    <voto scala="30">21</voto>
  </esame>
  <esame>
    <data>
      <mese>aprile</mese>
      <anno>2006</anno>
    </data>
    <voto scala="30">28</voto>
  </esame>
</risultati_esame>
```

Possibile Soluzione

```
<schema xmlns="..." xmlns:a="http://unirm3.it" targetNamespace="http://unirm3.it">

  <element name="risultati_esame">
    <complexType>
      <sequence>
        <element name="esame" type="a:tipo_esame" maxOccurs="unbounded"/>
      </sequence>
    </complexType>
  </element>

  <complexType name="tipo_esame">
    <sequence>
      <element name="data" type="a:tipo_data"/>
      <element name="voto" type="a:tipo_voto"/>
    </sequence>
  </complexType>

  <complexType name="tipo_data">
    <sequence>
      <element name="giorno" type="integer" minOccurs="0"/>
      <element name="mese" type="string"/>
      <element name="anno" type="integer"/>
    </sequence>
  </complexType>

  <complexType name="tipo_voto">
    <simpleContent>
      <extension base="integer">
        <attribute name="scala" type="integer"/>
      </extension>
    </simpleContent>
  </complexType>

</schema>
```

Domanda 3 (30%)

Considerare il seguente documento XML:

```
<?xml version="1.0" encoding="UTF-8"?>
<cineteca>
  <film anno="1994">
    <titolo>Blade Runner</titolo>
    <regista><cognome>Scott</cognome><nome>Ridley</nome></regista>
    <attori>
      <attore>Harrison Ford</attore>
      <attore>Rutger Hauer</attore>
    </attori>
    <produzione>Warner Bros</produzione>
    <durata>118</durata>
  </film>
  <film anno="1995">
    <titolo>Pulp Fiction</titolo>
    <regista><cognome>Tarantino</cognome><nome>Quentin</nome></regista>
    <attori>
      <attore>John Travolta</attore>
      <attore>Samuel L. Jackson</attore>
      <attore>Uma Thurman</attore>
    </attori>
    <produzione>Warner Bros</produzione>
    <durata>154</durata>
  </film>
  <film anno="1968">
    <titolo>2001: Odissea Nello Spazio</titolo>
    <regista><cognome>Kubrick</cognome><nome>Stanley</nome></regista>
    <attori>
      <attore>HAL 9000</attore>
    </attori>
    <produzione>Istituto Luce</produzione>
    <durata>141</durata>
  </film>
</cineteca>
```

Con riferimento a documenti XML di questi genere:

- (a) scrivere l'espressione XPATH che restituisce il primo attore di film precedenti al 1994
- (b) scrivere il foglio di stile XSLT che restituisce un file HTML che viene visualizzato come segue:

- **Blade Runner** - regista: RidleyScott
- **Pulp Fiction** - regista: QuentinTarantino
- **2001: Odissea Nello Spazio** - regista: StanleyKubrick

- (c) scrivere l'interrogazione XQuery che, per ogni attore, restituisce un elemento scheda-attore contenente il nome dell'attore e tutti i titoli dei film nei quali ha lavorato

Possibile Soluzione

(a)

```
doc("Cinema.xml")/cineteca/film[@anno<1994]/attori/attore[1]
```

(b)

```
<xsl:stylesheet ...>
  <xsl:template match="cineteca">
    <html>
      <head>
        <title>Films</title>
      </head>
      <body>
        <ul>
          <xsl:apply-templates select="film"/>
        </ul>
      </body>
    </html>
  </xsl:template>

  <xsl:template match="film">
    <li>
      <b><xsl:value-of select="titolo"/></b> -
      <xsl:apply-templates select="regista"/>
    </li>
  </xsl:template>

  <xsl:template match="regista">
    regista:
      <xsl:value-of select="nome"/>
      <xsl:value-of select="cognome"/>
  </xsl:template>
</xsl:stylesheet>
```

(c)

```
<attori>
{
  let $c := doc("Cinema.xml")/cineteca
  for $a in distinct-values($c//attore)
  return
  <scheda-attore >
    <nome-attore>{ $a }</nome-attore>
    {
      for $f in $c/film
      where $f//attore = $a
      return <film>{$f/titolo}</film>
    }
  </scheda-attore>
}
</attori>
```

Domanda 4 (20%)

Si consideri lo schema di relazione (che, nella realtà italiana, modella l'organizzazione delle zone postali nelle grandi città):
 $R(\text{Località, Via, CAP})$

con le dipendenze funzionali:

$$\begin{aligned} \text{CAP} &\rightarrow \text{Località} \\ \text{Località Via} &\rightarrow \text{CAP} \end{aligned}$$

- Giustificare la seguente affermazione: “lo schema R non è in BCNF e non esiste per esso alcuna buona decomposizione in BCNF”
- Modificare lo schema, con l'aggiunta di attributi o modifica di dipendenze funzionali, ottenendo uno schema per il quale sia possibile una buona decomposizione in BCNF (Suggerimento: una località può essere suddivisa in zone, ognuna delle quali corrisponde ad un certo CAP). Chiarire il significato dei nuovi attributi, mostrando anche un'istanza della nuova relazione
- Definire uno schema Entity-Relationship che descriva la realtà di interesse., traducendolo poi nel modello relazionale.

Possibile Soluzione

(Nota bene : lo schema in questo esercizio è analogo a quello “problematico” discusso a lezione e presentato anche nel testo di “Basi di dati”)

- Lo schema non è in BCNF perché la prima FD ha un primo membro (CAP) che non è superchiave (N.B. allo scopo, non è necessario trovare tutte le chiavi, ma, semplicemente, verificare, per ciascuna FD non banale, se il primo membro è superchiave, calcolandone la chiusura rispetto all'insieme di FD dato. Non può esistere una sua decomposizione adeguata in BCNF perché la seconda FD coinvolge tutti gli attributi e quindi una relazione su tutti gli attributi è necessaria per la conservazione delle dipendenze
- Si può introdurre un attributo “NumZona” che, insieme alla località, identifichi la una zona di CAP. Le FD diventano

$$\begin{aligned} \text{CAP} &\rightarrow \text{Località NumZona} \\ \text{Località NumZona} &\rightarrow \text{CAP} \\ \text{Località Via} &\rightarrow \text{Zona} \end{aligned}$$

Una possibile relazione è la seguente:

Località	NumZona	CAP	Via
Roma	46	00146	Via Vasca Navale
Roma	46	00146	Via Segre
Roma	85	00185	Viale delle Scienze
Milano	46	20146	Via Romagnoli

La relazione non è in BCNF, perché né CAP né LocalitàZona sono superchiavi, ma può essere decomposta in due relazioni

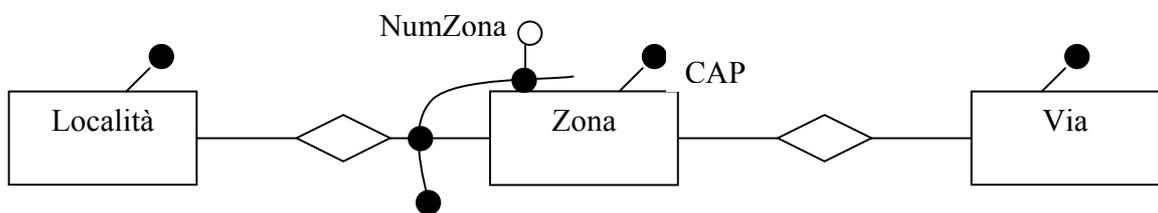
$$\begin{aligned} R1(\text{Località, Zona, CAP}) &\text{ con } \text{CAP} \rightarrow \text{Località} ; \text{CAP} \rightarrow \text{NumZona} ; \text{Località NumZona} \rightarrow \text{CAP} \\ R2(\text{Località, Via, NumZona}) &\text{ con } \text{Località Via} \rightarrow \text{NumZona} \end{aligned}$$

ciascuna delle quali è in BCNF. Una decomposizione dell'istanza è la seguente:

Località	NumZona	CAP
Roma	46	00146
Roma	85	00185
Milano	46	20146

Località	NumZona	Via
Roma	46	Via Vasca Navale
Roma	46	Via Segre
Roma	85	Viale delle Scienze
Milano	46	Via Romagnoli

- Uno schema ER che (con i soli attributi citati) modella la realtà di interesse è il seguente



Domanda 5 (20%)

Considerare il vincolo di integrità, detto dipendenza multivalore (MVD), indicato con $X \twoheadrightarrow Y$ e soddisfatto da una relazione $r(XYZ)$ se:

- per ogni coppia di ennuple t_1 e t_2 di r con $t_1[X]=t_2[X]$ si ha che r contiene una ennupla t con $t[XY]=t_1[XY]$ e $t[XZ]=t_2[XZ]$

Dimostrare che:

- se r soddisfa la FD $X \rightarrow Y$ allora r soddisfa anche la MVD $X \twoheadrightarrow Y$
- r soddisfa la MVD $X \twoheadrightarrow Y$ se e solo se r soddisfa la MVD $X \twoheadrightarrow Z$
- r si decompone senza perdita su $r_1(XY)$ e $r_2(XZ)$ se e solo se r soddisfa la MVD $X \twoheadrightarrow Y$

Possibile Soluzione

- se r soddisfa la FD $X \rightarrow Y$ allora r soddisfa anche la MVD $X \twoheadrightarrow Y$

Supponiamo che r soddisfi la FD $X \rightarrow Y$

Per dimostrare che r soddisfa anche la MVD $X \twoheadrightarrow Y$, consideriamo una qualunque coppia di ennuple t_1 e t_2 di r con $t_1[X]=t_2[X]$ e mostriamo che r contiene una ennupla t con $t[XY]=t_1[XY]$ e $t[XZ]=t_2[XZ]$.

In effetti, tale ennupla è proprio t_2 , perché essa è ovviamente uguale a se stessa su XZ e, poiché uguale a t_1 su X , risulta uguale a t_1 anche su Y , grazie alla FD $X \rightarrow Y$

- r soddisfa la MVD $X \twoheadrightarrow Y$ se e solo se r soddisfa la MVD $X \twoheadrightarrow Z$

Immediato, perché nella definizione Y e Z sono interscambiabili

- r si decompone senza perdita su $r_1(XY)$ e $r_2(XZ)$ se e solo se r soddisfa la MVD $X \twoheadrightarrow Y$
 - “se” (se r soddisfa $X \twoheadrightarrow Y$ allora si decompone senza perdita)
Consideriamo una ennupla t nel join delle proiezioni. Per definizione degli operatori di join e proiezione, essa proviene da una ennupla nella proiezione su XY e da una ennupla nella proiezione su XZ e quindi da due ennuple t_1 e t_2 con $t_1[XY]=t[XY]$ e $t_2[XZ]=t[XZ]$; poiché r soddisfa la MVD $X \twoheadrightarrow Y$, allora per definizione di MVD, abbiamo che t appartiene a r
 - “solo se” (se r si decompone senza perdita allora soddisfa $X \twoheadrightarrow Y$)
Consideriamo una ogni coppia di ennuple t_1 e t_2 di r con $t_1[X]=t_2[X]$; il join delle proiezioni, per definizione degli operatori, contiene una ennupla t con $t[XY]=t_1[XY]$ e $t[XZ]=t_2[XZ]$; poiché la relazione si decompone senza perdita, allora tale ennupla appartiene anche a r ed è quindi la ennupla che ci garantisce che t_1 e t_2 non causano la violazione della MVD.

(Nota bene: questa proprietà non può essere dimostrata facendo riferimento al teorema sulla decomposizione senza perdita rispetto a FD, perché non si ha nessuna informazione sulle FD)