

Information Integration – Exercise 1 – Gabriel Glaser

Task 1: Passing from an XML Schema to a relational schema

a)

```
CREATE TABLE ShipOrders (  
  orderID int(255) NOT NULL,  
  orderPerson varchar(100),  
  shiptoName varchar(100),  
  shiptoAddress varchar(100),  
  shipToCity varchar(100),  
  shipToCountry varchar(100),  
  PRIMARY KEY (orderID)  
);
```

```
CREATE TABLE Items (  
  itemID int(255),  
  itemTitle varchar(100),  
  itemNote varchar(1000),  
  itemQuantity int(255) CHECK (item_quantity >= 1),  
  itemPrice decimal(6, 2),  
  orderID int(255),  
  PRIMARY KEY (itemID)  
  FOREIGN KEY (orderID) REFERENCES ShipOrders(orderID)  
);
```

b) Constraints:

- orderPerson, shiptoName, shiptoAddress, shipToCity, shipToCountry boundable ✓
- orderID required ✓
- More item entries possible, but ≥ 1 is not enforced (can be 0 item entries) ✗
- Variable itemNote optional (could be set to **NULL** in table, if not present in XML) ✓
- Variable quantity positive ✓
- Variable price decimal ✓

c) SQL requires size definitions (*varchar* and *decimal*), which are not present in the XML schema.

Task 2: Querying an integrated system

a) Example web form for integrated system:

CDTitle	
releaseYear	<input type="button" value="exact"/> <input type="button" value="before"/> <input type="button" value="after"/>
trackTitle	
name	
position	<input type="button" value="exact"/> <input type="button" value="before"/> <input type="button" value="after"/>
language ▼	
German	
French	
Spanish	
...	

Search

b) Which queries can be answered by the integrating system?

- **Q1: Yes.** *CDTitle* bound ✓, only use of *AND* ✓, *language* bound ✓, no bind of not explicitly allowed variables ✓ and no comparisons with other operators than =, < or > ✓.
- **Q2: No.** *language* not bound.
- **Q3: No.** None of *CDTitle*, *name*, *trackTitle* bound.
- **Q4: No.** None of *CDTitle*, *name*, *trackTitle* bound and use of illegal operator on *position* variable.
- **Q5: No.** Uses *OR* but only *AND* is allowed.
- **Q6: Yes.** *CDTitle* bound ✓, only use of *AND* ✓, *language* bound ✓, no bind of not explicitly allowed variables ✓ and only valid operators ✓.

c) Transformation of legal, global queries to each source format:

- *Q1*:

- *Amazon*:

KeyWords = ""
Artist = ""
Title = "The Endless River"
Label = ""
EAN/UPC = ""

(*language* not possible to bind)

- *MusicBrainz*:

SELECT * FROM Release
WHERE Release.release = "The Endless River" AND
Release.lang="EN"

- *Q6*:

- *Amazon*:

KeyWords = ""
Artist = "Taylor Swift"
Title = "1989"
Label = ""
EAN/UPC = ""

(*language, releaseYear* not possible to bind)

- *MusicBrainz*:

SELECT * FROM Artist, Release
WHERE Artist.artist="Taylor Swift" AND
Release.release = "1989" AND
Release.date>"2010-12-31" AND
Release.lang="EN"