Zawansowane Modelowanie
i Analiza Systemów
Informatycznych
(l- 5)

Polsko-Japońska Wyższa Szkoła Technik Komputerowych
Katedra Systemów Informacyjnych
2013
Overview

• Step 6: Add value, set comparison and subtyping constraints (cont)

• Step 7; final checks

• Some observations on schema transformations

• Go to orm.net for the readings
The 7 steps of CSDP

Step 1: transform familiar information examples into elementary facts, and apply quality checks
Step 2: draw the fact types, and apply a population check
Step 3: check for entity types that should be combined, and note any arithmetic derivations
Step 4: add uniqueness constraints, arity of fact types, splitting of fact types.
Step 5: add mandatory role constraints, and check for logical derivations
Step 6: add value, set comparison and subtyping constraints
Step 7: add other constraints and perform final quality checks (e.g., populating fact type instances)
### Step 6 - A subtype construction one more example: Library Acquisition System

<table>
<thead>
<tr>
<th>ACCESSION#</th>
<th>ISBN</th>
<th>CATALOGUE</th>
<th>TYPE</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>23456</td>
<td>0-8053-0145-3</td>
<td>530.03</td>
<td>monograph</td>
<td>1989</td>
</tr>
<tr>
<td>23457</td>
<td>0-7248-0151-0</td>
<td>530.03</td>
<td>monograph</td>
<td>1989</td>
</tr>
<tr>
<td>65473</td>
<td>-</td>
<td>531.02</td>
<td>serial</td>
<td>Dec 1987</td>
</tr>
<tr>
<td>75436</td>
<td>-</td>
<td>531.02</td>
<td>serial</td>
<td>Jan 1988</td>
</tr>
<tr>
<td>31873</td>
<td>-</td>
<td>400.02</td>
<td>video</td>
<td>1976</td>
</tr>
<tr>
<td>54326</td>
<td>-</td>
<td>994.04</td>
<td>software</td>
<td>1986</td>
</tr>
<tr>
<td>87965</td>
<td>-</td>
<td>530.03</td>
<td>serial</td>
<td>Dec 1989</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACCESSION#</th>
<th>AUTHOR</th>
<th>EDITOR</th>
<th>TITLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>23456</td>
<td>R Elmasri</td>
<td>-</td>
<td>Fundamentals of Database Systems</td>
</tr>
<tr>
<td></td>
<td>S Navathe</td>
<td>-</td>
<td>Conceptual Schema and Relational Database Design</td>
</tr>
<tr>
<td></td>
<td>G M Nijssen</td>
<td>-</td>
<td>Communications of the ACM</td>
</tr>
<tr>
<td></td>
<td>T A Halpin</td>
<td>-</td>
<td>Language at Twelve</td>
</tr>
<tr>
<td>23457</td>
<td>-</td>
<td>P Denning</td>
<td>Communications of the ACM</td>
</tr>
<tr>
<td>65473</td>
<td>-</td>
<td>P Denning</td>
<td>Histobank: The 1930's Depression</td>
</tr>
<tr>
<td>75436</td>
<td>-</td>
<td>-</td>
<td>Transactions on Database Systems</td>
</tr>
<tr>
<td>31873</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>54326</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>87965</td>
<td>-</td>
<td>G Wiederhold</td>
<td></td>
</tr>
<tr>
<td>ACCESSION#</td>
<td>VOLUME</td>
<td>NUMBER</td>
<td>FORMAT</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td>23456</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>23457</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>65473</td>
<td>30</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>75436</td>
<td>31</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>31873</td>
<td>-</td>
<td>-</td>
<td>VHS</td>
</tr>
<tr>
<td>54326</td>
<td>-</td>
<td>-</td>
<td>Apple</td>
</tr>
<tr>
<td>87965</td>
<td>14</td>
<td>4</td>
<td>-</td>
</tr>
</tbody>
</table>
If we look at the output report, there are various kinds of holding

*monograph*
*serial*
*video*
*software*

Intuitively, one may suggest the following split of ‘Library Item’ into four subtypes:
Redrawing the diagram produces

Note:
This schema should be refined. One can notice that there are two fact types ‘Copies’. It suggests that two designated subtypes involved may have something in common. A formal way of subtype construction gives different result.

ZMA-5
To analyse the output report from the UoD consider the rows of information provided in the output report and the roles recorded.

The following is the object/role table

<table>
<thead>
<tr>
<th>has access number</th>
<th>publish in year</th>
<th>publish in month</th>
<th>has title</th>
<th>has catalog as</th>
<th>has type</th>
<th>has ISBN</th>
<th>has format with copies</th>
<th>runs on</th>
<th>copies held</th>
<th>has volume#</th>
<th>has serial#</th>
<th>edited by</th>
<th>written by</th>
</tr>
</thead>
<tbody>
<tr>
<td>23456</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>23457</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>65473</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>75436</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>31873</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>54326</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>87965</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

The table only records the presence of a fact.
The table can be reorganised to more clearly show the pattern of facts:

<table>
<thead>
<tr>
<th>has access number</th>
<th>publish in year</th>
<th>has title</th>
<th>catalog as</th>
<th>has type</th>
<th>publish in month</th>
<th>has volume#</th>
<th>has serial#</th>
<th>edited by</th>
<th>copies held</th>
<th>has ISBN</th>
<th>written by</th>
<th>has format /with copies</th>
<th>runs on</th>
</tr>
</thead>
<tbody>
<tr>
<td>23456</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23457</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65473</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75436</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31873</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>54326</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>87965</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Various rows are identical. They can be collapsed into single rows. Identical columns can be grouped together to form the subtyping structure.

<table>
<thead>
<tr>
<th>Group</th>
<th>has access number</th>
<th>publish in year</th>
<th>has title</th>
<th>catalog as</th>
<th>has type</th>
<th>publish in month</th>
<th>has volume#</th>
<th>has serial#</th>
<th>edited by</th>
<th>copies held</th>
<th>has ISBN</th>
<th>written by</th>
<th>has format /with copies</th>
<th>runs on</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ZMA-5
Recall, that if for every ‘tick’ in the column representing set X there exists a ‘tick’ in the same rows in a column representing the set Y, then there is a subtype relationship between X and Y, more precisely X is a subtype of Y (notation X $\rightarrow$ Y).

**In our example**

<table>
<thead>
<tr>
<th>Group</th>
<th>has access number</th>
<th>publish in year</th>
<th>has title</th>
<th>catalog as</th>
<th>has type</th>
<th>publish in month</th>
<th>has volume#</th>
<th>has serial#</th>
<th>edited by</th>
<th>copies held</th>
<th>has ISBN</th>
<th>written by</th>
<th>Has format</th>
<th>runs on</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>/with copies</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>/with copies</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>/with copies</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>/with copies</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$B \rightarrow A$

$E \rightarrow A$

$F \rightarrow A$

$D \rightarrow C$

$D \rightarrow A$
B \rightarrow A, C \rightarrow A, D \rightarrow A, E \rightarrow A, F \rightarrow A, D \rightarrow C, F \rightarrow C.

Subtype graph:  

After removal of redundant links:

Subtype graph structure with names added:

Library item

Serial

Non_serial

Software

Video

Monograph
Subtyped conceptual schema diagram
(a) Disjoint and not total
S1 ∩ S2 = ∅
S1 ∪ S2 ≠ S

Example
S - All employees
S1- Middle management
S2 - Factory workers

(b) Disjoint and total
S1 ∩ S2 = ∅
S1 ∪ S2 = S

Example
S - All Students
S1- Post Graduate
S2 - Under Graduate
(a) Not disjoint and not total
\[ S_1 \cap S_2 \neq \emptyset \]
\[ S_1 \cup S_2 \text{ included in } S \]

Example
S - All students
S1 - Postgraduate Students
S2 - Female Students

(b) Not disjoint and total
\[ S_1 \cap S_2 \neq \emptyset \]
\[ S_1 \cup S_2 = S \]

Example
S - All university people
S1 - Students
S2 - University employees
Correct schema
Library example
Consider another example Universe ‘Vehicles’ described by the following data sample

<table>
<thead>
<tr>
<th>TRUCKS</th>
<th>V_no</th>
<th>Price</th>
<th>WheelNo</th>
<th>WheelSz</th>
<th>FuelCons</th>
<th>MaxLoad</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>20000</td>
<td>6</td>
<td>40</td>
<td>30</td>
<td>5000</td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>40000</td>
<td>4</td>
<td>40</td>
<td>15</td>
<td>500</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BOATS</th>
<th>V_no</th>
<th>Price</th>
<th>NoSeats</th>
<th>EngPower</th>
</tr>
</thead>
<tbody>
<tr>
<td>1002</td>
<td>100000</td>
<td>6</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>1005</td>
<td>80000</td>
<td>4</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>1007</td>
<td>40000</td>
<td>4</td>
<td>150</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CARS</th>
<th>V_no</th>
<th>Price</th>
<th>WheelNo</th>
<th>WheelSz</th>
<th>FuelCons</th>
<th>MaxSpeed</th>
<th>NoSeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>1003</td>
<td>18000</td>
<td>4</td>
<td>25</td>
<td>12</td>
<td>160</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BIKES</th>
<th>V_no</th>
<th>Price</th>
<th>WheelNo</th>
<th>WheelSz</th>
<th>NoSeats</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1004</td>
<td>200</td>
<td>2</td>
<td>50</td>
<td>2</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>
To analyze the output report from the UoD consider the rows of information provided in the output report and the roles recorded.

The following table (matrix) is used to analyze the UoD ‘Vehicles’. We are interested in the existence of the property recorded, rather than in its value.

<table>
<thead>
<tr>
<th>V_no</th>
<th>Prce</th>
<th>WhNo</th>
<th>WhSz</th>
<th>Fuel</th>
<th>MaxLd</th>
<th>Seats</th>
<th>EngPow</th>
<th>MaxSp</th>
<th>Wght</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1002</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1005</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1007</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1003</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1004</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ZMA-5
Various rows can be identical. They can be collapsed into single rows. Identical columns should be grouped together to form the subtyping subsets.

<table>
<thead>
<tr>
<th>Gr_no</th>
<th>Price</th>
<th>WhNo</th>
<th>WhSz</th>
<th>Fuel</th>
<th>MaxLd</th>
<th>Seats</th>
<th>EngPow</th>
<th>MaxSp</th>
<th>Wght</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>.</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
</tbody>
</table>

Such constructed matrix is a source of information for the construction of the subtype structure. Each column ‘pattern’ corresponds to one subtype and a relationship between column patterns is used to determine the relationship between the corresponding subtypes.
Recall what we have defined before - If for every tick/cross in the column representing a set X there exists a ‘tick’ in the same rows in a column representing the set Y, then we say that there is a subtype relationship between X and Y, more precisely X is a subtype of Y (notation $X \rightarrow Y$).

- **X** is a subtype of **Y**.
  - **X** → **Y**
  - Each cross in column X has a corresponding cross on the same level in the Y column.

- **X** is not a subtype of **Y**.
  - **Y** is not a subtype of **X**
  - The cross in the second row in column X has no corresponding cross on the same level in the column Y.
  - Crosses in the fifth and sixth row in the column Y have no corresponding crosses in the column X.
In our example:

<table>
<thead>
<tr>
<th>G_no</th>
<th>Prce</th>
<th>WhNo</th>
<th>WhSz</th>
<th>Fuel</th>
<th>MaxLd</th>
<th>Seats</th>
<th>EngPow</th>
<th>MaxSp</th>
<th>Wght</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

. A  B  B  C  D  E  F  G  H

All are subtypes of A.

C \rightarrow B, D \rightarrow B, D \rightarrow C, F \rightarrow E, G \rightarrow E, H \rightarrow E, G \rightarrow C, H \rightarrow B, G \rightarrow B
Some of the subtype relationships are redundant. This follows from the transitivity of the subset relation. Redundant subtypes are marked in red.

The nonredundant subtypes that generate subtype graph are;

\[ \begin{align*}
B &\rightarrow A, E \rightarrow A, C \rightarrow B, \\
D &\rightarrow C, F \rightarrow E, G \rightarrow E, \\
H &\rightarrow E, G \rightarrow C, H \rightarrow B, \\
\end{align*} \]
How to name the subtypes?
We try to generate the names using semantics of the roles that generate them, unless there is another property that defines the subtype:

A - All vehicles
B - Vehicles with wheels
C - Fuel consuming vehicles
D - Trucks
E - Passenger Vehicles
F - Boats
G - Cars
H - Bicycles
Interesting subtype construction application

Subtype analysis can be useful for the classification of responses to wide range of input forms that contain a number of conditional instructions (for instance tax return, application forms or surveys).

We present an example of a subtype construction for the responses based on the structure of conditional instructions in a questionnaire.

Note that typically, every respondent must answer the first question. Instructions that follow answer options (a, b, etc) direct the particular respondent to another question (not necessarily the subsequent one) by specifying the question number or simply terminate the questionnaire.

Let us consider the following example
Q1. What is your main occupation
   a) Academic working in CS 2
   b) CS Industry prof 2
   c) Academic or industry prof from other fields 10

Q2. How much leisure time you have for yourself for sport and other non-work activities (estimate average per day):
   a) close to 0 than you
   b) 20 3
   c) 40 3
   d) 60 or more minutes 3

Q3. What are your favorite activities
   a) Doing nothing (this includes watching TV) 5
   b) Hobby linked to your profession 5
   c) Hobby not linked to your profession 4

Q4. What is your main hobby (choose one option)
   a) Cooking 5
   b) Gardening 5
   c) woodworking 5
   d) metalworking 5
   e) other, please specify…….. 5

Q5. How much time you devote to sport activities
   a) 0 than you
   b) a lot, but only watching sport on TV than you
   c) 30 6
   d) 60 (or 60+) min per day 6
Q6. What kind of sport you mainly do
   a) playing games (like tennis, soccer, etc) 7
   b) track and field type 8
   c) other, please specify …………… 9

Q7. Your partners are mainly
   a) CS people 9
   b) Business people 9
   c) You care only about their sporting skills 9

Q8. How do you evaluate your fitness
   a) very good 9
   b) good 9
   c) average 9
   d) disaster 9

Q9. If you have been asked to answer Q4 please answer this one as well. Otherwise go straight to 10.
What impact on your work does your active leisure have?
   a) Strong 10
   b) Neutral 10
   c) Negative 10
   d) no opinion 10

Q10. Do you consider a fitness of CS professionals as
   a) above society average
   b) average
   c) below average
   d) no opinion
Construction of the questionnaire flow

- Each question is represented by a token and has connection with another token if corresponding question is present in one of its options.
- For example; 6 will be connected to 7, 8 and 9 since these question are choices from options of the question 6 as the next questions.
- For termination option a special symbol is to be used (see Q2 and Q5)
- If there is some not typical instruction included in the question then mark it with question mark and give it a consideration in the step of determination of respondent patterns
Navigation in the graph is essential. One has to make sure that all the paths will be identified. There are different approaches to such searching. We use preorder traversal.

Here we only present a graphical “illustration’ of such method.
Discovery of different respondent patterns in the answering questionnaire. Proper navigation in the graph is essential and very important.

1, 10
1, 2
1, 2, 3, 4, 5
1, 2, 3, 4, 5, 6, 9, 10
1, 2, 3, 4, 5, 6, 7, 9, 10
1, 2, 3, 4, 5, 6, 8, 9, 10
1, 2, 3, 5
1, 2, 3, 5, 6, 10 (??: no ‘4’ in the path so no ‘9’)
1, 2, 3, 5, 6, 7, 10 (??: no ‘4’ so no ‘9’)
1, 2, 3, 5, 6, 8, 10 (??: no ‘4’ so no ‘9’)

ZMA-5
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>F</th>
<th>C</th>
<th>E</th>
<th>H</th>
<th>I</th>
<th>G</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>8</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>9</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1, 10  
1, 2  
1, 2, 3, 4, 5  
1, 2, 3, 4, 5, 6, 9, 10  
1, 2, 3, 4, 5, 6, 7, 9, 10  
1, 2, 3, 4, 5, 6, 8, 9, 10  
1, 2, 3, 5  
1, 2, 3, 5, 6, 10  
1, 2, 3, 5, 6, 7, 10  
1, 2, 3, 5, 6, 8, 10
Now we process this matrix as the ordinary object/role matrices.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>F</td>
<td>C</td>
<td>E</td>
<td>H</td>
<td>I</td>
<td>G</td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

All subsets of A
B → A
C → B
D → A
E → D, E → C
F → C
G → E, G → F
H → E
I → E
Note that each subtype corresponds to a question (or a set of questions). Naming of subtypes here is based on the property in the answer option(s) that directs the respondent to that question (or to the question with the smallest number, in case of a set of questions defining the subtype).

For example: Subtype C is generated by Q3 and Q5. Options in Q2 that direct respondent to Q 3 are selected by people who spent some time doing sport and other activities (NonLazy 😊).
Step 7 – Other Constraints

Entity type (label type) constraints

Two types of constraint are shown:
- set cardinality; number of allowable instances of an entity type
- allowable values of a label type

The purpose - show on the schema that some standards have to be enforced

Example 1:

```
Employee (EMP#) -----> Classification
                  |
                  v
Position (P_CODE) <->

P_CODE: {prog, prog/anal, sys anal, dba, sys prog, EUC coord}

card(Employee) < 100
```

Example 2:

```
Day -----> Day_designation
          |
          v

Day_Name: {Mon, Tue, Wed, Thu, Fri, Sat, Sun}
```

Example 3:

```
Employee (EMP#) -----> Home -----> Address (ADDR)
                        |
                        v
EMP#: {100000..999999}
ADDR: char(60)
```
Step 7 - Role value constraints

A role may be allowed to hold only for a subset of the values allowed for its entity.

In the following example Money allows any value permitted by label type real(8,2), whereas Salary can only hold values of Money between 20000 and 80000 and Expenses can be between 500 and 5000. This information can be used in defining triggers in database application.

![Diagram](chart.png)
**Step 7 - Add fact type occurrence frequency constraints and set comparisons**

Fact type occurrence: A student may enrol a minimum of 1 subject and a maximum of 5 subjects.

Comparisons: A student can’t enrol a subject and take leave at the same time.

How about the number of subjects lectured by a lecturer greater than the number of publications???

ORM does not provide a possibility to show it in the above schema
Add other constraints and perform final checks.
• check if each entity can be identified.
• check the conceptual schema can hold the prescribed data
• check missing constraints e.g. a subject can only be enrolled by students if there is a lecturer to hold that subject.
Step 7 (con’t)

Can the fact type population satisfy all the constraints specified in your conceptual schema? Can all the population (data) in the UoD be captured by your conceptual schema?

Count(student) <= 1000

Sub-code:[ITB001..ITN555]

f/t lecturer

p/t lecturer

Sub-code: [ITB001..ITB999]

02244556 ITB220
02244557 ITB225
02244557 ITB256
02244557 ITB106
02244557 ITB220

02244559 L001
02244559 L002
01944559 L003

02244559 L001
02244559 L002
01944559 L003

R.Lau P001
J. Reye P002
K.Ling P003

On-leave

leave (leave-id)

enrolment

Sub-code: [ITB001..ITB999]

02244556 ITB220
02244557 ITB225
02244557 ITB256
02244557 ITB106
02244557 ITB220

02244559 L001
02244559 L002
01944559 L003

S641 J. Reye
S540 K.Ling

Office (off#)

Off- taken

Has-off

subject

(sub-code)

lecture
d-by

lecture
d-by

lecturer

(name)

publication (publication-id)

research

conducted

Performed-by

ZMA-5
Occurrence Frequencies – more formally

A set of occurrence frequencies are defined upon a role, or combination of roles, within a fact type.

It specifies how often an instance of an entity type, or combination of entity types, may be involved in a role, or combination of roles.

One or both limits in occurrence frequency constraints does not have to be constant – could vary upon some properties of entity types involved in the role affected by that constraint.
Occurrence Frequencies – for a single role

\[ E \rightarrow R \rightarrow A \]

Occurrence Frequencies – for multiple roles

\[ B \rightarrow C \leftarrow F \rightarrow A \]

\[ B \quad C \quad A \]

ZMA-5
Example

The conceptual schema diagram states that a student must be enrolled in between 1 and 5 units.

<table>
<thead>
<tr>
<th>Stu#</th>
<th>Unit_code</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1</td>
<td>ITB220</td>
</tr>
<tr>
<td>s1</td>
<td>ITB222</td>
</tr>
<tr>
<td>s1</td>
<td>ITB223</td>
</tr>
<tr>
<td>s3</td>
<td>ITB220</td>
</tr>
<tr>
<td>s3</td>
<td>ITB220</td>
</tr>
<tr>
<td>s4</td>
<td>ITB222</td>
</tr>
<tr>
<td>s4</td>
<td>ITB222</td>
</tr>
<tr>
<td>s5</td>
<td>ITB220</td>
</tr>
<tr>
<td>s5</td>
<td>ITB220</td>
</tr>
<tr>
<td>s6</td>
<td>ITB222</td>
</tr>
<tr>
<td>s6</td>
<td>ITB222</td>
</tr>
<tr>
<td>s7</td>
<td>ITB220</td>
</tr>
<tr>
<td>s7</td>
<td>ITB220</td>
</tr>
<tr>
<td>s8</td>
<td>ITB220</td>
</tr>
<tr>
<td>s8</td>
<td>ITB222</td>
</tr>
<tr>
<td>s9</td>
<td>ITB220</td>
</tr>
<tr>
<td>s9</td>
<td>ITB222</td>
</tr>
<tr>
<td>s10</td>
<td>ITB220</td>
</tr>
<tr>
<td>s10</td>
<td>ITB222</td>
</tr>
<tr>
<td>s11</td>
<td>ITB222</td>
</tr>
<tr>
<td>s12</td>
<td>ITB222</td>
</tr>
</tbody>
</table>
Example (cont) with more accurate information

The conceptual schema diagram should include the representation of the following constraints:
1. Students must be enrolled in between 1 and 5 units in a single semester,
2. Unit must attract enrolment of between 10 and 100 students to be run.

Note that the occurrence frequency relates to the combination of roles.
Population Check

Check if the conceptual schema has too strong constraint anomaly. i.e. is the output report, on which the conceptual schema is based, a permitted population of the conceptual schema?

This can be checked using the natural join operation on the populated fact types.
Check that each entity can be identified.
i.e. check that all instances of every entity type have a unique naming mechanism.

**Example** – block of land

**Example**
Example

Unique identification depends on the particular semantic limits of the UoD. The identification of a city house will be of different accuracy if UoD is restricted to one CITY or restricted to a Country or to a Continent.
Other constraints

Example:
Marital status

A single person changes to married
A married person changes to divorced or widowed
A divorced person changes to married
A widowed person changes to married.

A Legal_change table can be stored in the database and to be used in triggers for checking if updates of Marital status is consistent with legal rules.
## Conceptual Schema Transformation

Different "conceptual views" have the same semantics.

<table>
<thead>
<tr>
<th>Reductions</th>
<th>Compositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>splitting -</td>
<td>grouping -</td>
</tr>
<tr>
<td>nesting</td>
<td>flattening</td>
</tr>
<tr>
<td>entity type to role reduction</td>
<td>role to entity type composition</td>
</tr>
<tr>
<td>fact type to entity type conversion</td>
<td>entity type to fact type conversion</td>
</tr>
</tbody>
</table>
Flattening

Nesting

Office (OFFICE#)

Building (BUILDING#)

Area (M²)

Office_areas

has area

contains

is area of

Office (OFFICE#)

located in building

contains office

Building (BUILDING#)

Area (M²)

ZMA-5 47
Nesting
Sub-relationship objectified for schema ‘optimisation’
Entity type to role reduction:

Role to entity type composition
Entity type to role reduction

Different schemas for the same UoD.

Resulting in different structure of data storage.
Fact type to entity type conversion
Jumping transition

Preferred representation:

- **Person (P_NAME)**
- **Marriage_state**
- **MaritalStat (StatCode)**

**is single**

**is married**
Which is the "best" conceptual view?

General guidelines:

- avoid entity types which are nothing but surrogates for fact types;
- whenever in two or more fact types, the same sub-relationship occurs, objectify this sub-relationship;
- avoid jumping transitions. That is, where the old fact and the new fact are of different types.
Summary

We have covered the basic method of ORM conceptual schema construction.

To get skills in this method substantial practice is required.
Go to orm.net for the readings

• What's new?
• The ORM 2009 Workshop, November 4-6, 2009 in Vilamoura, Portugal.
• The ORM 2010 Workshop, October 27-29, 2010 in Hersonissou, Crete, Greece.
• The ORM 2011: International Workshop on Fact-Oriented Modeling at Hersonissou, Crete, Greece, October 2011
• The ORM 2012: International Workshop on Fact-Oriented Modeling Rome, Italy, September 12-14, 2012

http://www.ormfoundation.org