Zawansowane Modelowanie i Analiza Systemów Informatycznych

(l- 5)

Polsko-Japońska Wyższa Szkoła Technik Komputerowych
Katedra Systemów Informacyjnych
2013

Overwiev

- **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

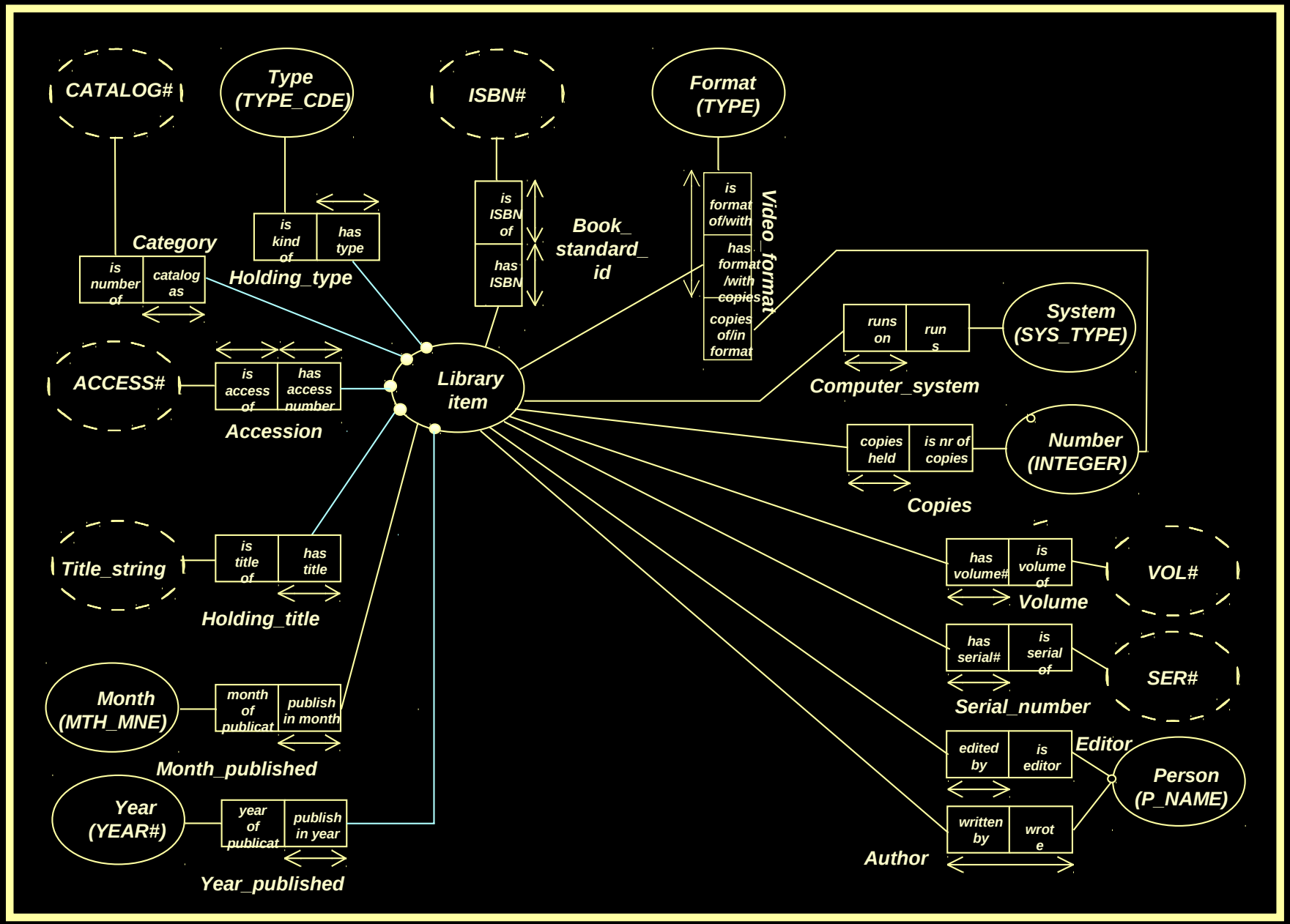
ACCESSION#	ISBN	CATALOGUE	TYPE	DATE
23456	0-8053-0145-3	530.03	monograph	1989
23457	0-7248-0151-0	530.03	monograph	1989
65473	-	531.02	serial	Dec 1987
75436	-	531.02	serial	Jan 1988
31873	-	400.02	video	1976
54326	-	994.04	software	1986
87965	-	530.03	serial	Dec 1989

ACCESSION#	AUTHOR	EDITOR	TITLE
23456	R Elmasri S Navathe	-	Fundamentals of Database Systems
23457	G M Nijssen T A Halpin	-	Conceptual Schema and Relational Database Design
65473	-	P Denning	Communications of the ACM
75436	-	P Denning	Communications of the ACM
31873	-	-	Language at Twelve
54326	-	-	Histobank: The 1930's Depression
87965	-	G Wiederhold	Transactions on Database Systems

Example: cont

ACCESSION#	VOLUME	NUMBER	FORMAT	SYSTEM	COPIES
23456	-	-	-	-	3
23457	-	-	-	-	2
65473	30	12	-	-	-
75436	31	1	-	-	-
31873	-	-	VHS U-matic	- -	3 1
54326	-	-	-	Apple	1
87965	14	4	-	-	

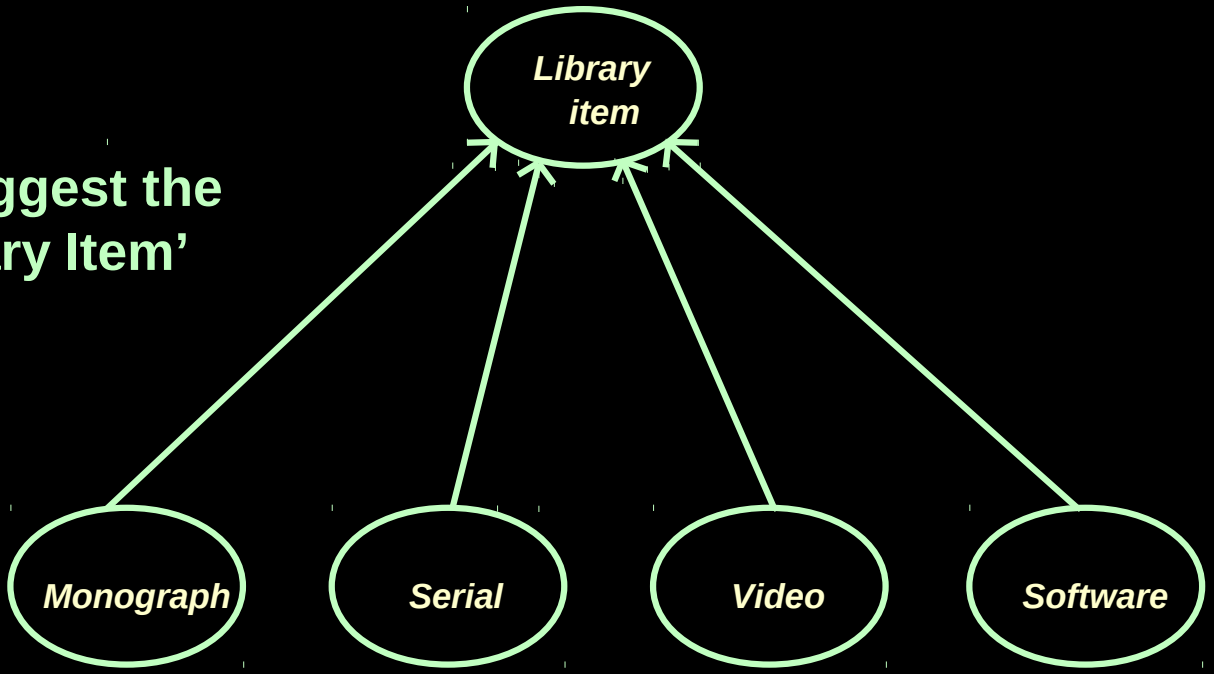
The conceptual schema diagram



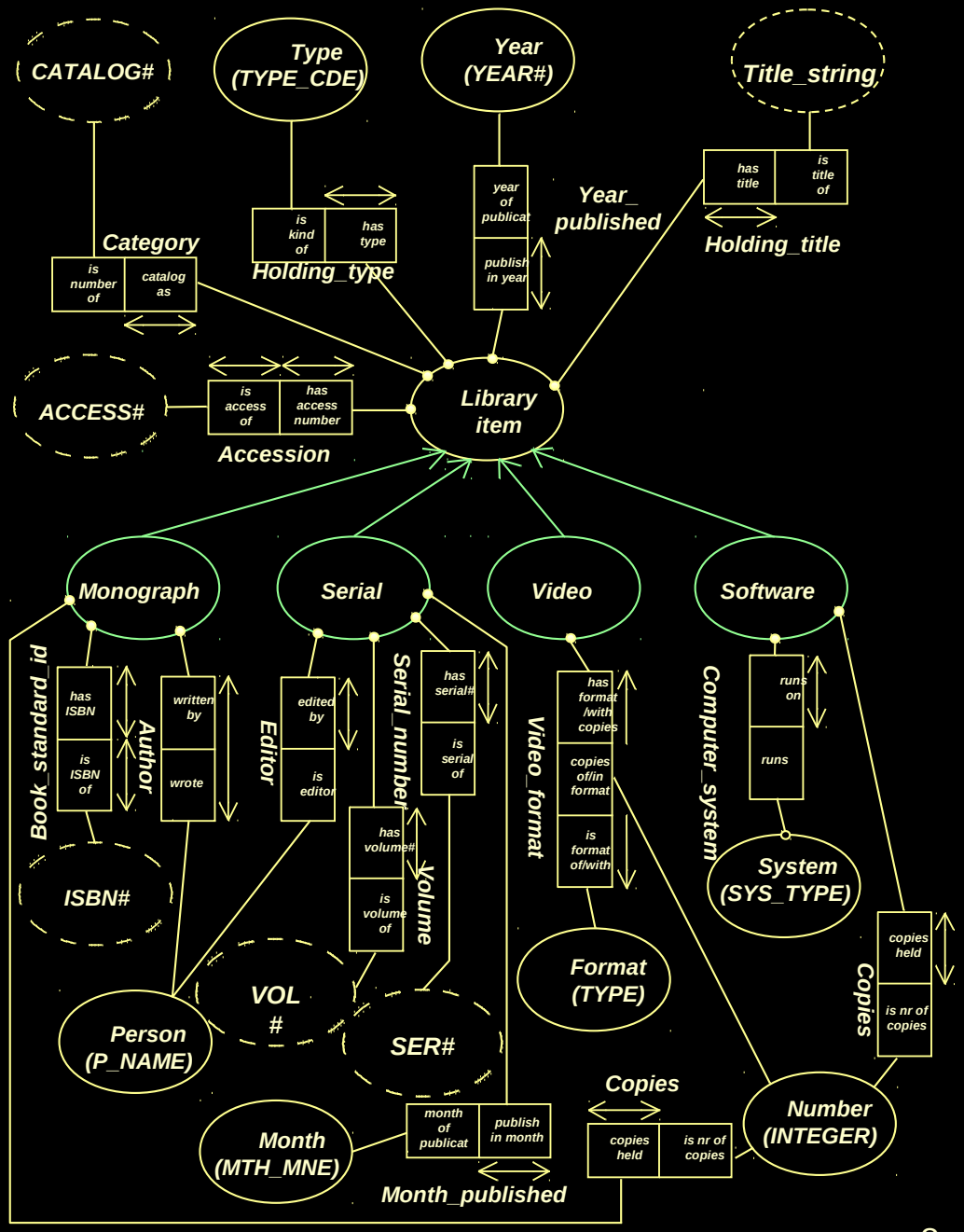
If we look at the output report, there are various kinds of holding

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.

To analyse the output report from the UoD consider the rows of information provided in the output report and the <u>roles</u> recorded.

The following is the object/role table

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

The table only records the presence of a fact.

The table can be reorganised to more clearly show the pattern of facts:

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

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

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

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→Y).

In our example

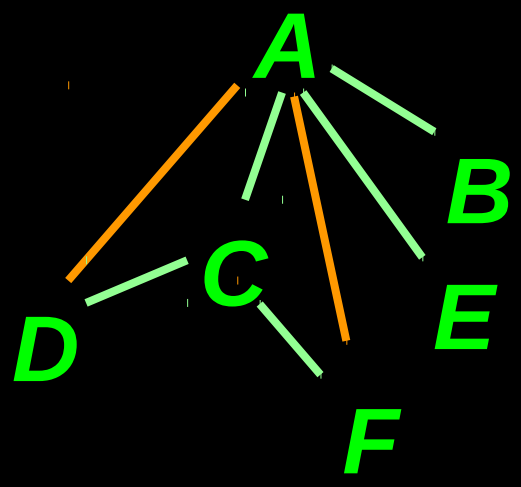
Group	has access number	publish in year	has title	catalog as	has type	publish in month	has volume#	has serial#	edited by	copies held	has ISBN	written by	Has format /with copies	runs on
1	✓	✓	✓	✓	✓					✓	✓	✓		
2	✓	✓	✓	✓	✓	✓	✓	✓	✓					
3	✓	✓	✓	✓	✓								✓	
4	✓	✓	✓	✓	✓					✓				✓
	A					B				C	D		E	F

B→A **C→A** **F→C** **D→C**

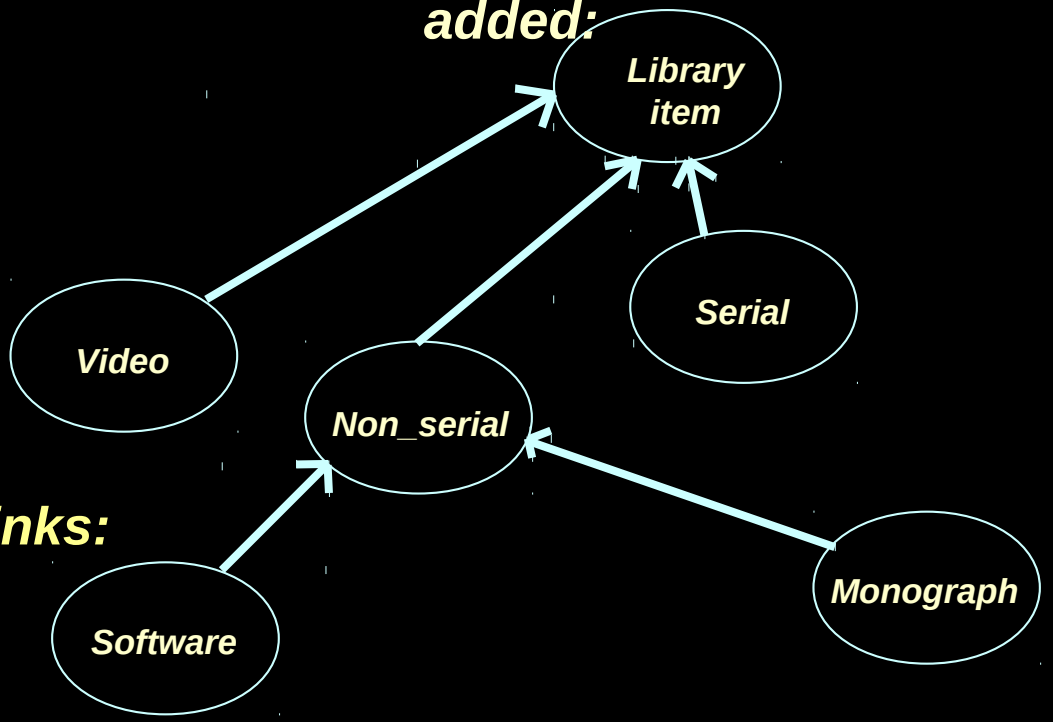
E→A **F→A** **D→A**

B→A, C→A, D→A, E→A, F→A, D→C, F→C.

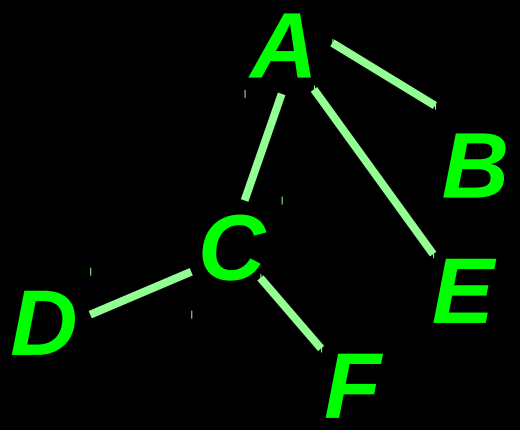
Subtype graph:



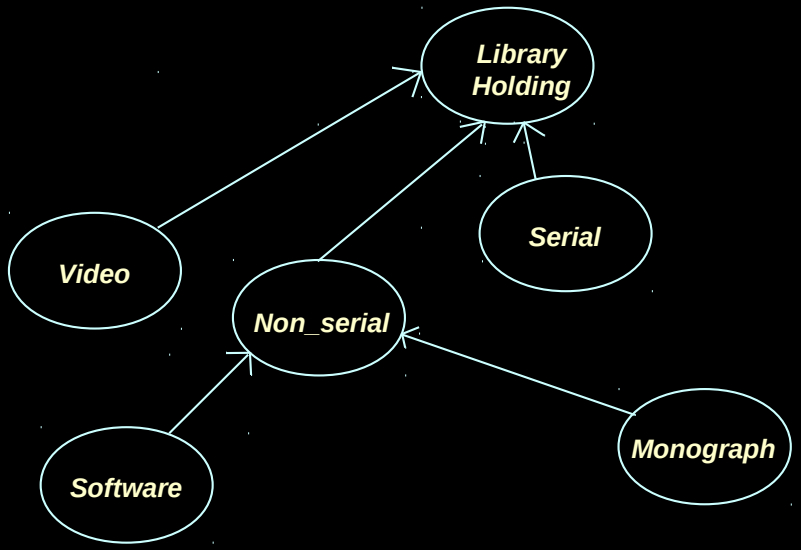
Subtype graph structure with names added:

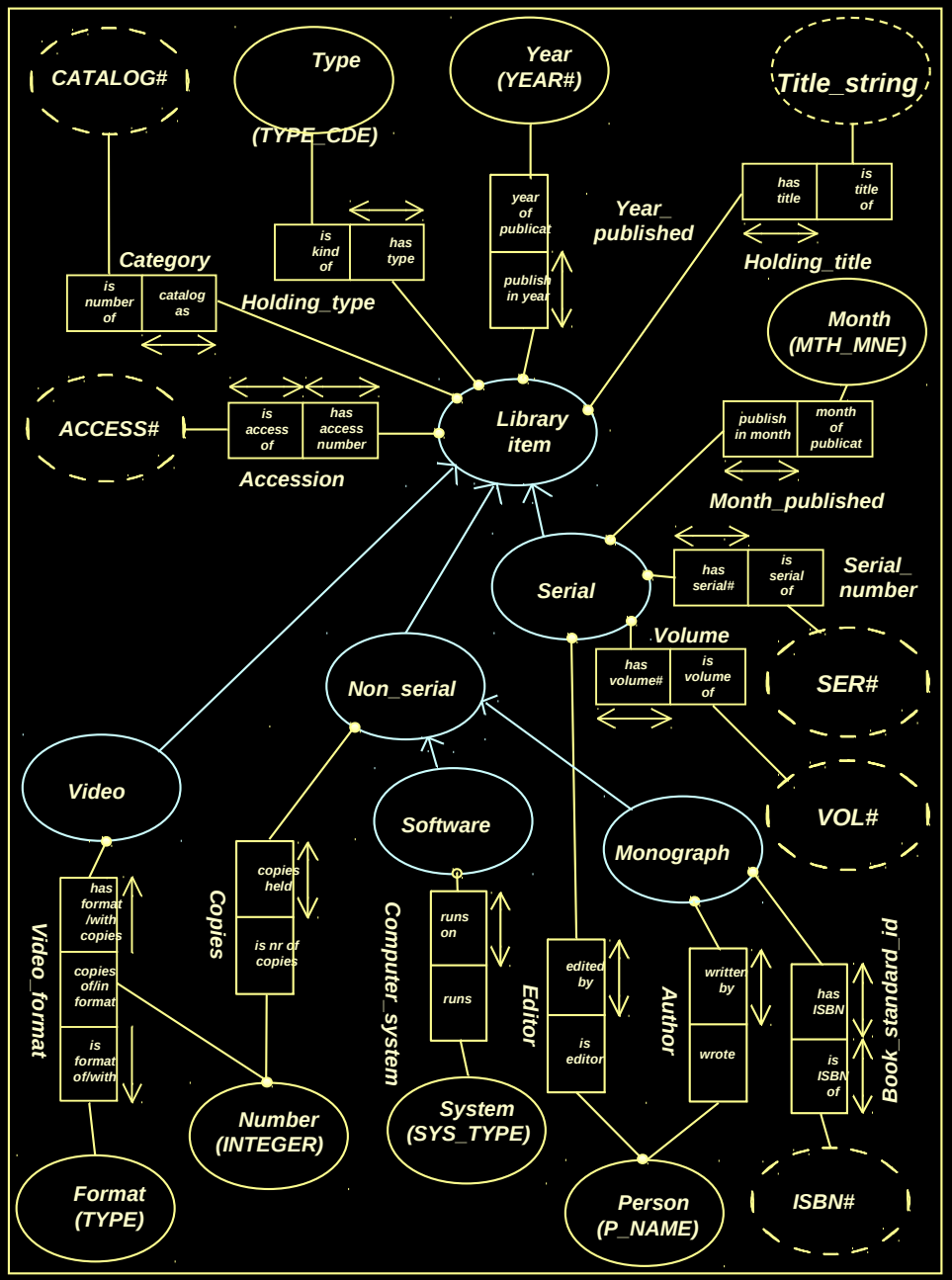


After removal of redundant links:



Subtyped conceptual schema diagram



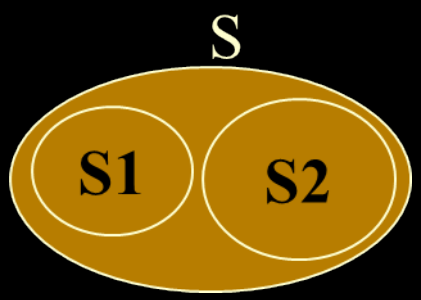


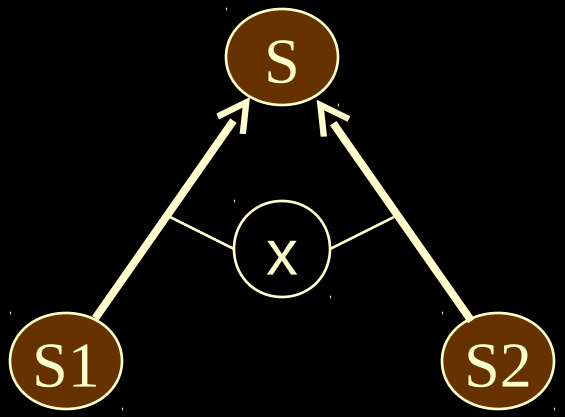
Subtype Notation

(a) Disjoint and not total

S1 ∩ S2 = ∅

S1 ∪ S2 ≠ S



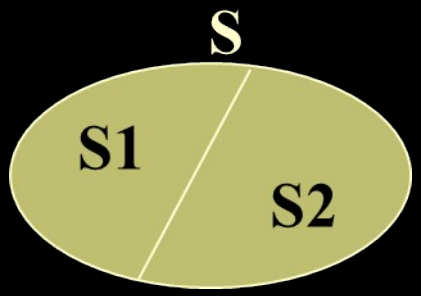


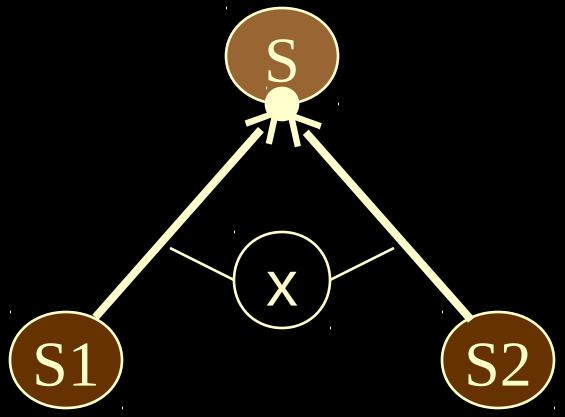
Example
S - All employees
S1- Midlle management
S2 - Factory workers

(b) Disjoint and total

S1 ∩ S2 = ∅

S1 ∪ S2 = S



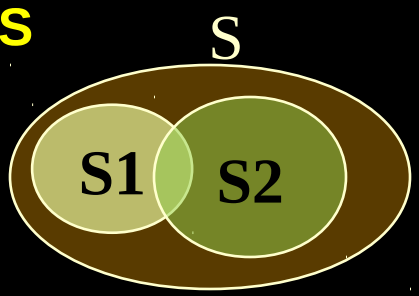


Example
S - All Students
S1- Post Graduate
S2 - Under Graduate

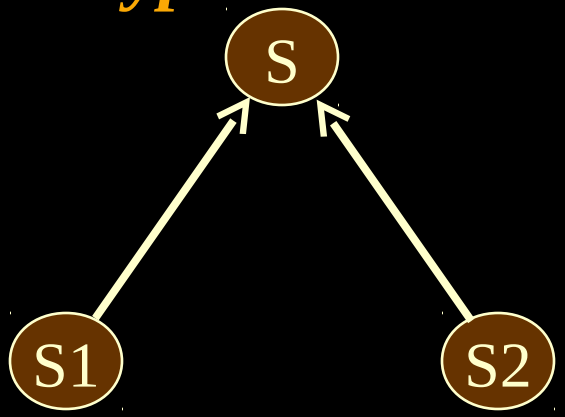
(a) Not disjoint and not total

$S1 \cap S2 \neq \emptyset$

$S1 \cup S2$ included in S



Subtype Notation

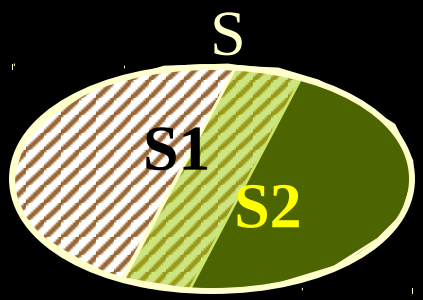


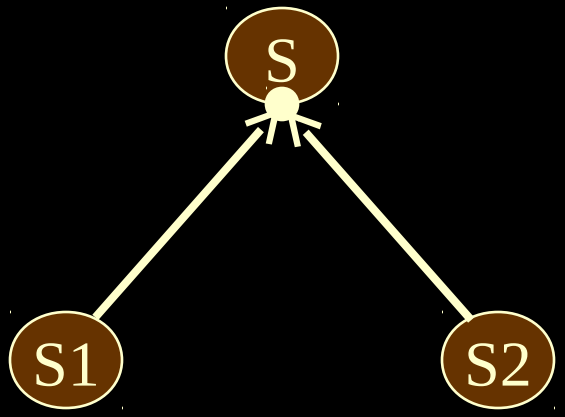
Example
S - All students
S1- Postgraduate Students
S2 - Female Students

(b) Not disjoint and total

$S1 \cap S2 \neq \emptyset$

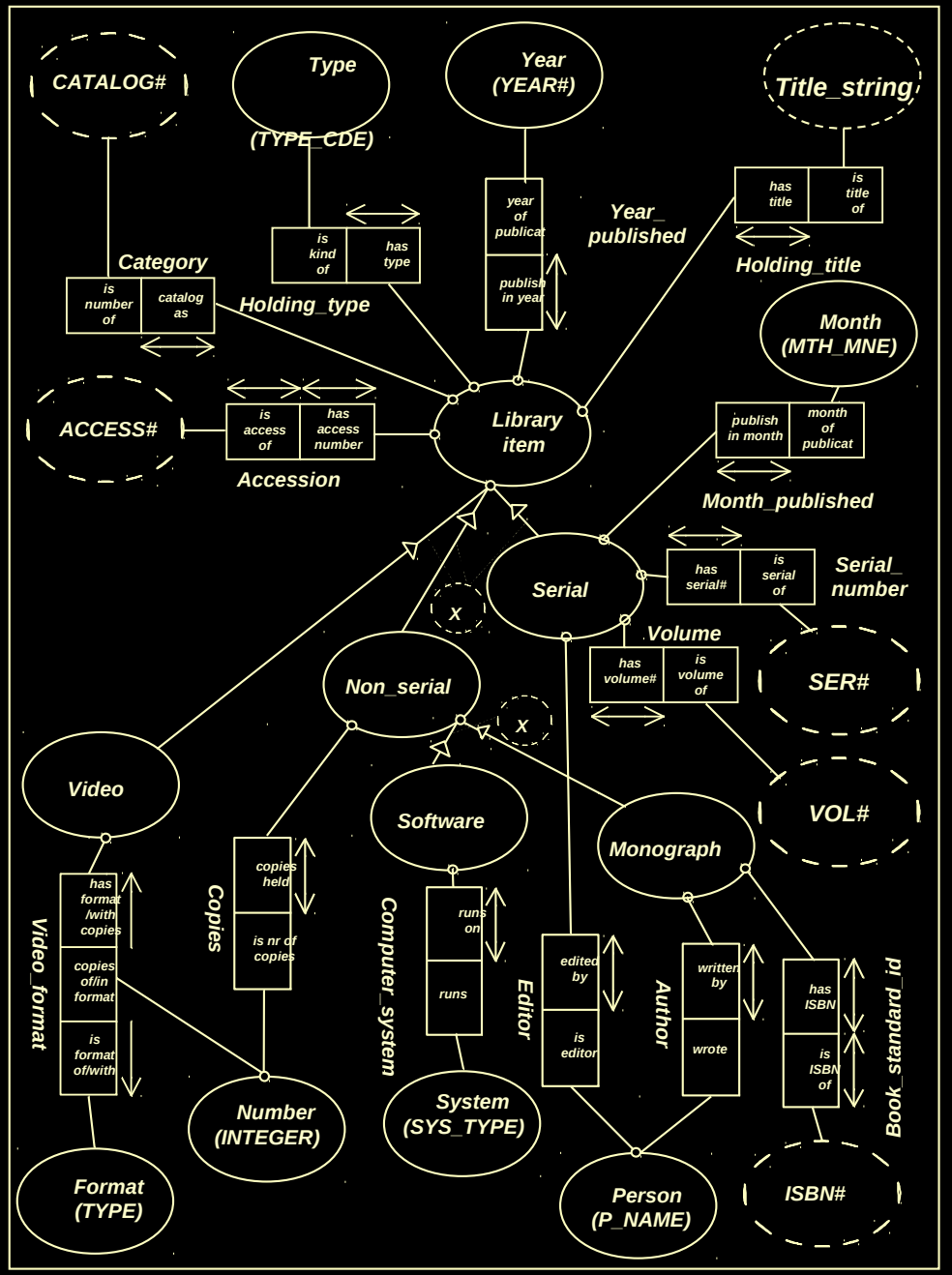
$S1 \cup S2 = S$





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

TRUCKS

V_no	Price	WheelNo	WheelSz	FuelCons	MaxLoad
1000	20000	6	40	30	5000
1001	40000	4	40	15	500

BOATS

V_no	Price	NoSeats	EngPower
1002	100000	6	200
1005	80000	4	150
1007	40000	4	150

CARS

V_no	Price	WheelNo	WheelSz	FuelCons	MaxSpeed	NoSeats
1003	18000	4	25	12	160	5

BIKES

V_no	Price	WheelNo	WheelSz	NoSeats	Weight
1004	200	2	50	2	25

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

V_no	Prce	WhNo	WhSz	Fuel	MaxLd	Seats	EngPow	MaxSp	Wght
1000	X	X	X	X	X				
1001	X	X	X	X	X				
1002	X					X	X		
1005	X					X	X		
1007	X					X	X		
1003	X	X	X	X		X		X	
1004	X	X	X			X			X

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

Gr_no	Price	WhNo	WhSz	Fuel	MaxLd	Seats	EngPow	MaxSp	Wght
1	X	X	X	X	X				
2	X					X	X		
3	X	X	X	X		X		X	
4	X	X	X			X			X
.	A	B	B	C	D	E	F	G	H

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→Y).

X	X
	X
X	X
	X
X	X
X	X
X	Y

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

X	X
X	
X	X
X	X
	X
	X
X	X
X	Y

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 a coresponding 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;

G_no	Prce	WhNo	WhSz	Fuel	MaxLd	Seats	EngPow	MaxSp	Wght
1	X	X	X	X	X				
2	X					X	X		
3	X	X	X	X		X		X	
4	X	X	X			X			X
.	A	B	B	C	D	E	F	G	H

All are subtypes of A.

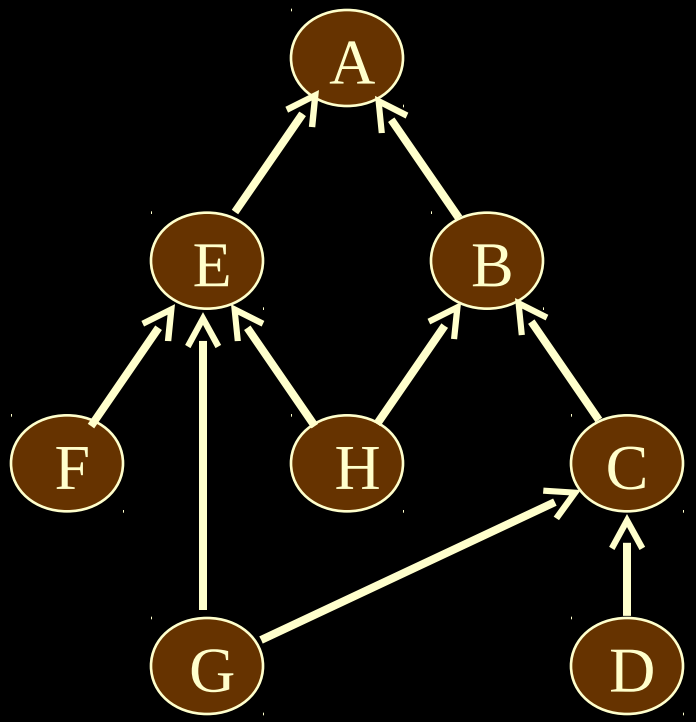
C → B, D → B, D → C, F → E, G → E, H → E, G → C, H → B, G → B

B → A, C → A, D → A, E → A, F → A, G → A, H → A,
C → B, D → B, D → C, F → E, G → E, H → E, G → C, H → B, G → 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;

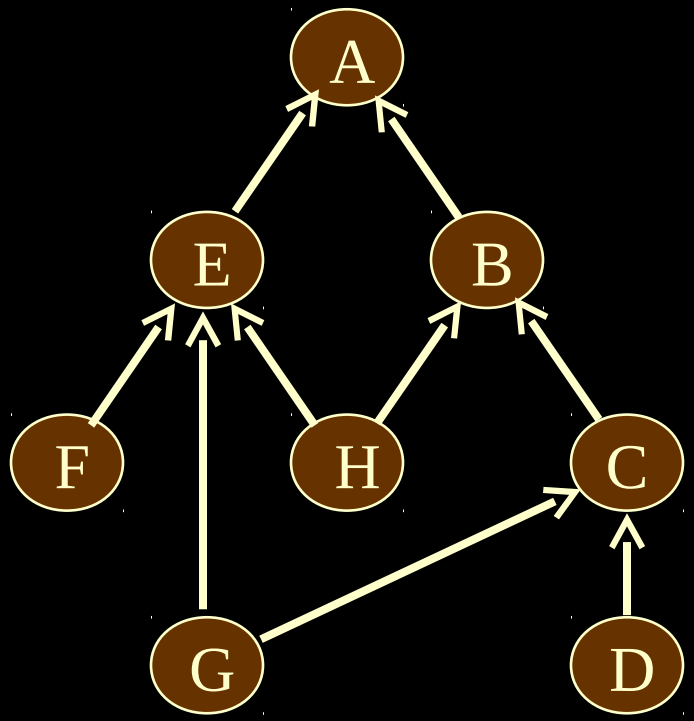
B → A, E → A, C → B,
D → C, F → E, G → E,
H → E, G → C, H → B,



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



G_no	Prce	WhNo	WhSz	Fuel	MaxLd	Seats	EngPow	MaxSp	Wght
1	X	X	X	X	X				
2	X					X	X		
3	X	X	X	X		X		X	
4	X	X	X			X			X
·	A	B	B	C	D	E	F	G	H

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 subseqent 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 perday);

a) close to 0	thank 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	thank you
b) a lot, but only watching sport on TV	thank 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

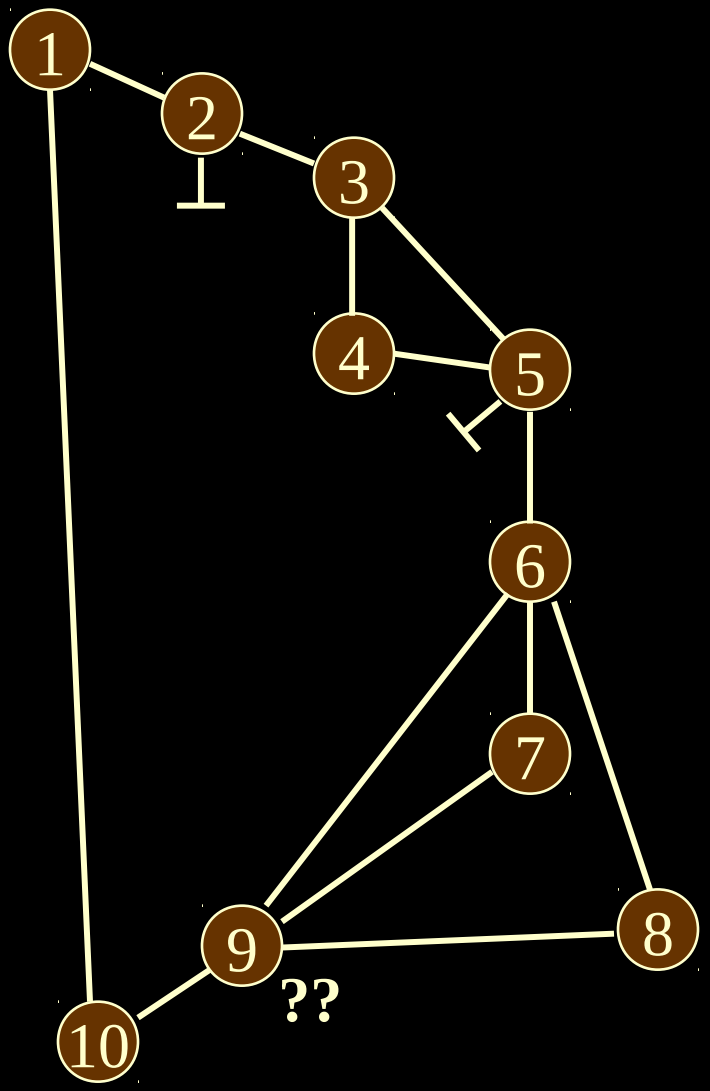
a) very good	9
b) good	9
c) average	9
d) disaster	9

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

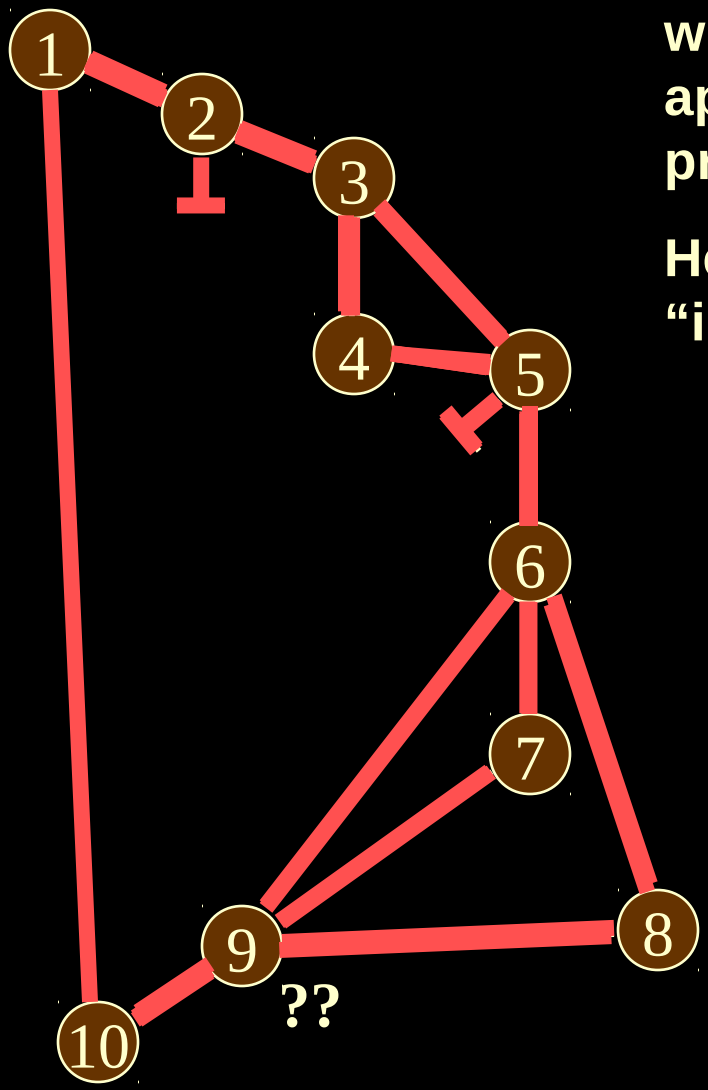
Q10. Do you consider a fitness of CS professionals as

a) above society average
b) average
c) below average
d) no opinion



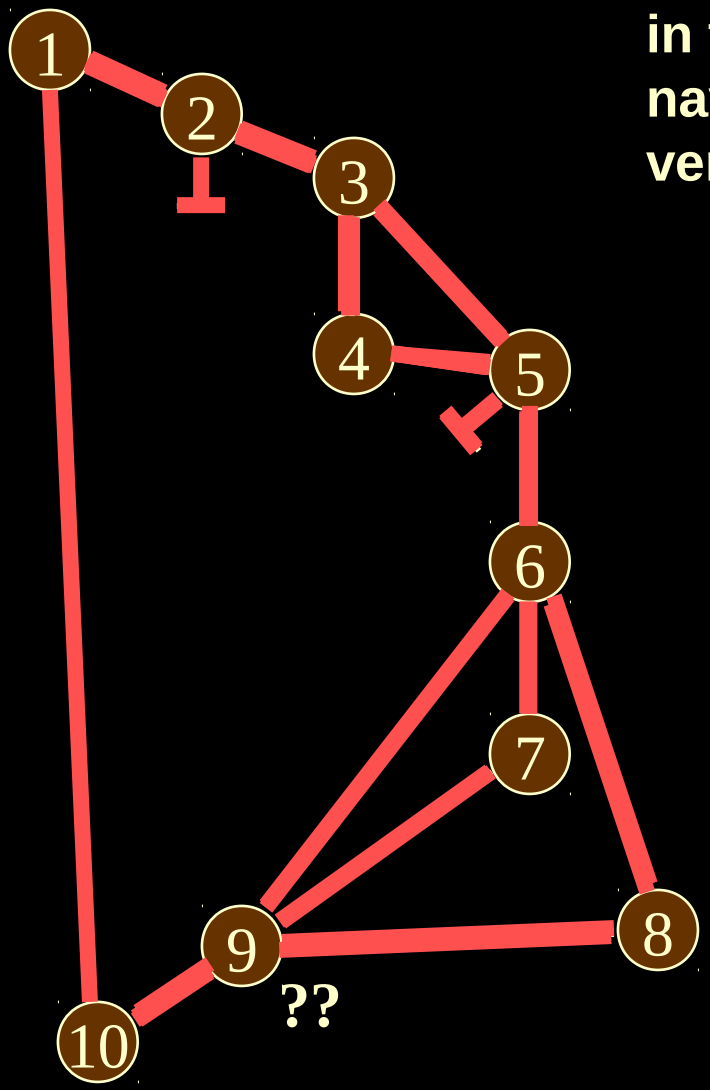
Construction of the questionnaire flow

- 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’)

1	2	3	4	5	6	7	8	9	10	
X									X	1, 10
X	X									1, 2
X	X	X	X	X						1, 2, 3, 4, 5
X	X	X	X	X	X			X	X	1, 2, 3, 4, 5, 6, 9, 10
X	X	X	X	X	X	X		X	X	1, 2, 3, 4, 5, 6, 7, 9, 10
X	X	X	X	X	X		X	X	X	1, 2, 3, 4, 5, 6, 8, 9, 10
X	X	X		X						1, 2, 3, 5
X	X	X		X	X				X	1, 2, 3, 5, 6, 10
X	X	X		X	X	X			X	1, 2, 3, 5, 6, 7, 10
X	X	X		X	X		X		X	1, 2, 3, 5, 6, 8, 10
A	B	C	F	C	E	H	I	G	D	

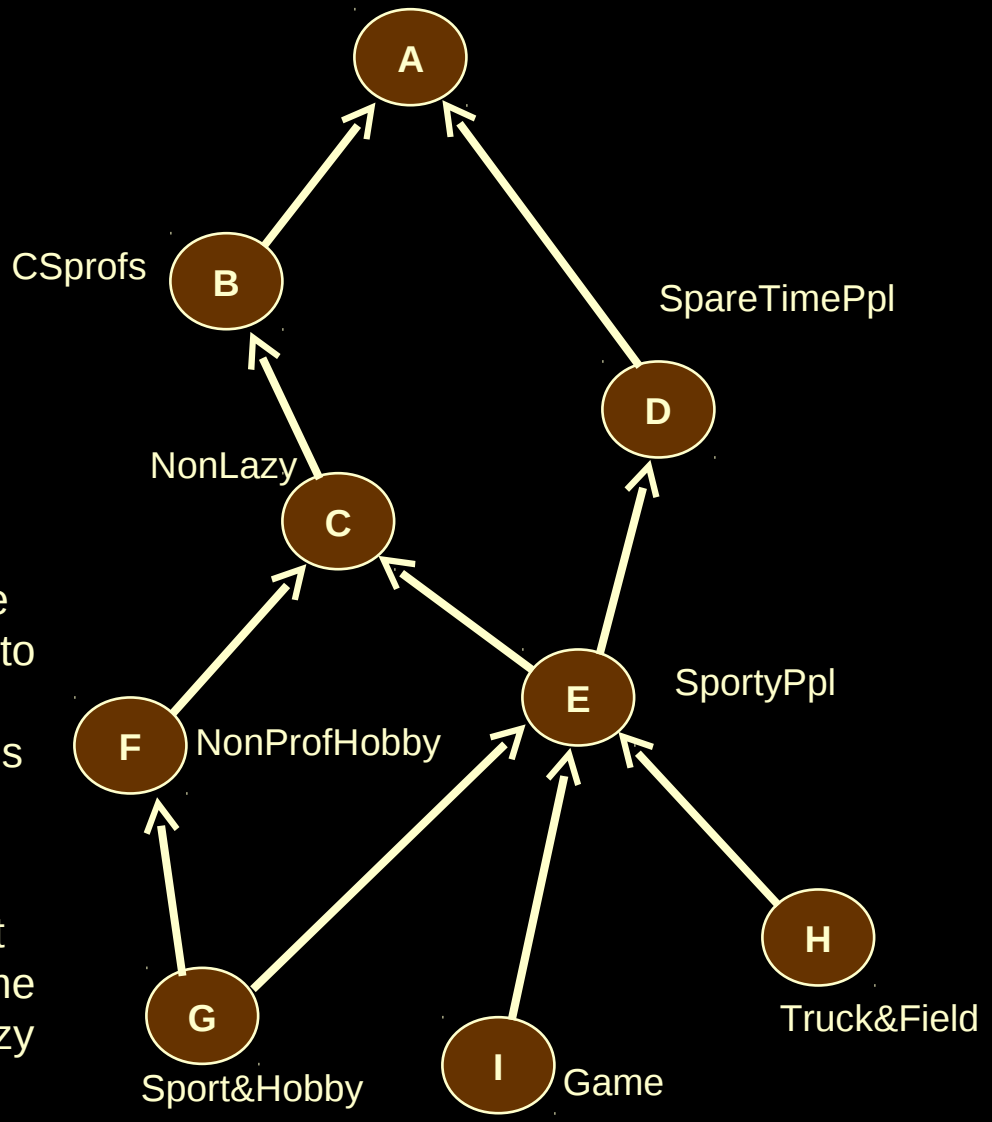
Now we process this matrics as the ordinary object/role matrics.

1	2	3	4	5	6	7	8	9	10
X									X
X	X								
X	X	X	X	X					
X	X	X	X	X	X			X	X
X	X	X	X	X	X	X		X	X
X	X	X	X	X	X		X	X	X
X	X	X		X					
X	X	X		X	X				X
X	X	X		X	X	X			X
X	X	X		X	X		X		X
A	B	C	F	C	E	H	I	G	D

All subsets of A

B → A
C → B
D → A
E → D , E → C
F → C
G → E, G→F
H → E
I → E

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

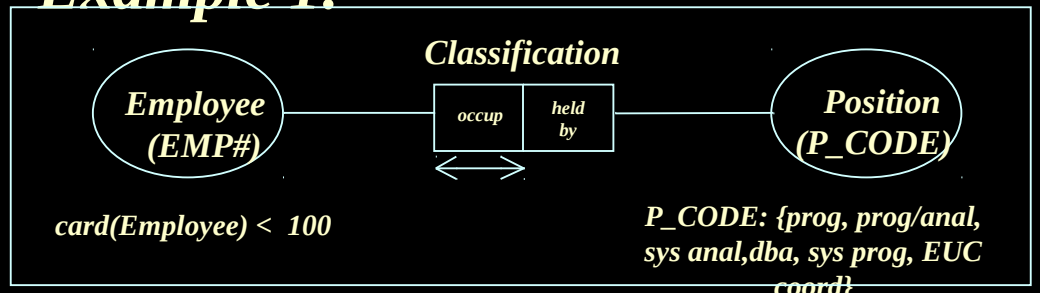
Entity type (label type) constraints

Two types of constraint are shown:

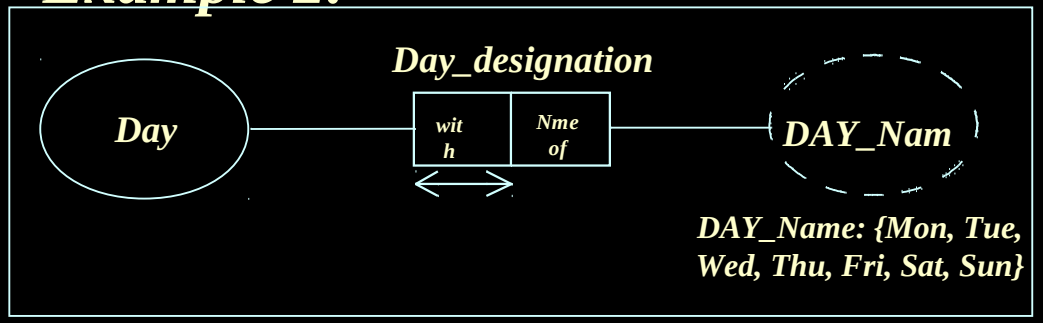
- **set cardinality; number of allowable instances of an <u>entity type</u>**
- **allowable values of a <u>label type</u>**

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

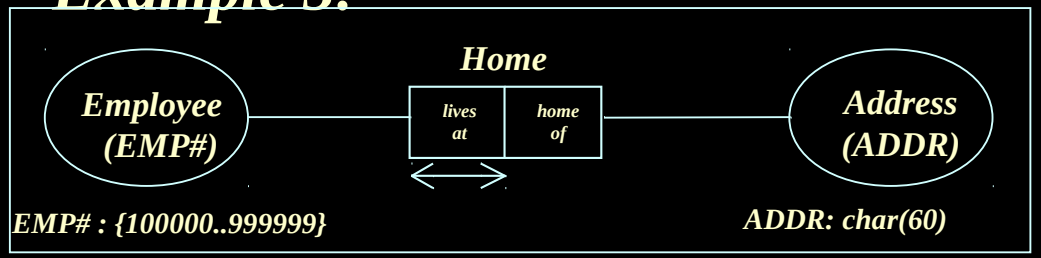
Example 1:



Example 2:

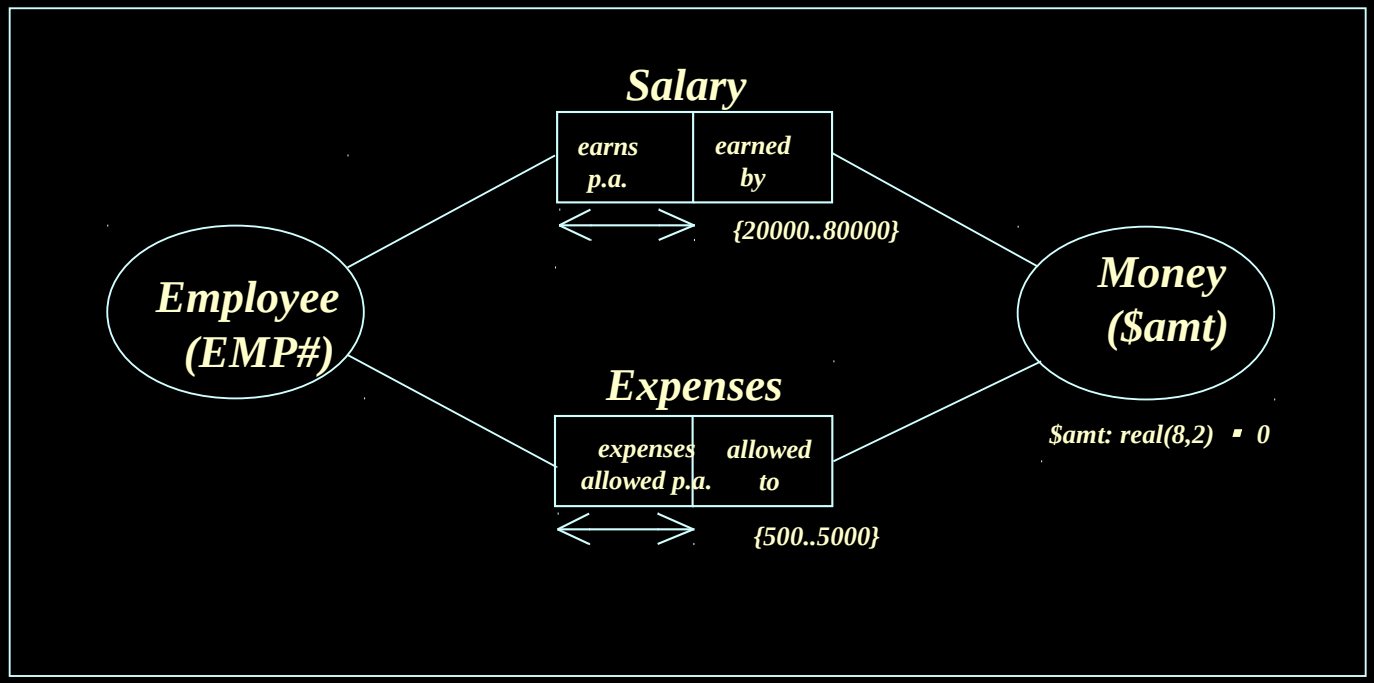


Example 3:



Step 7 - Role value constraints

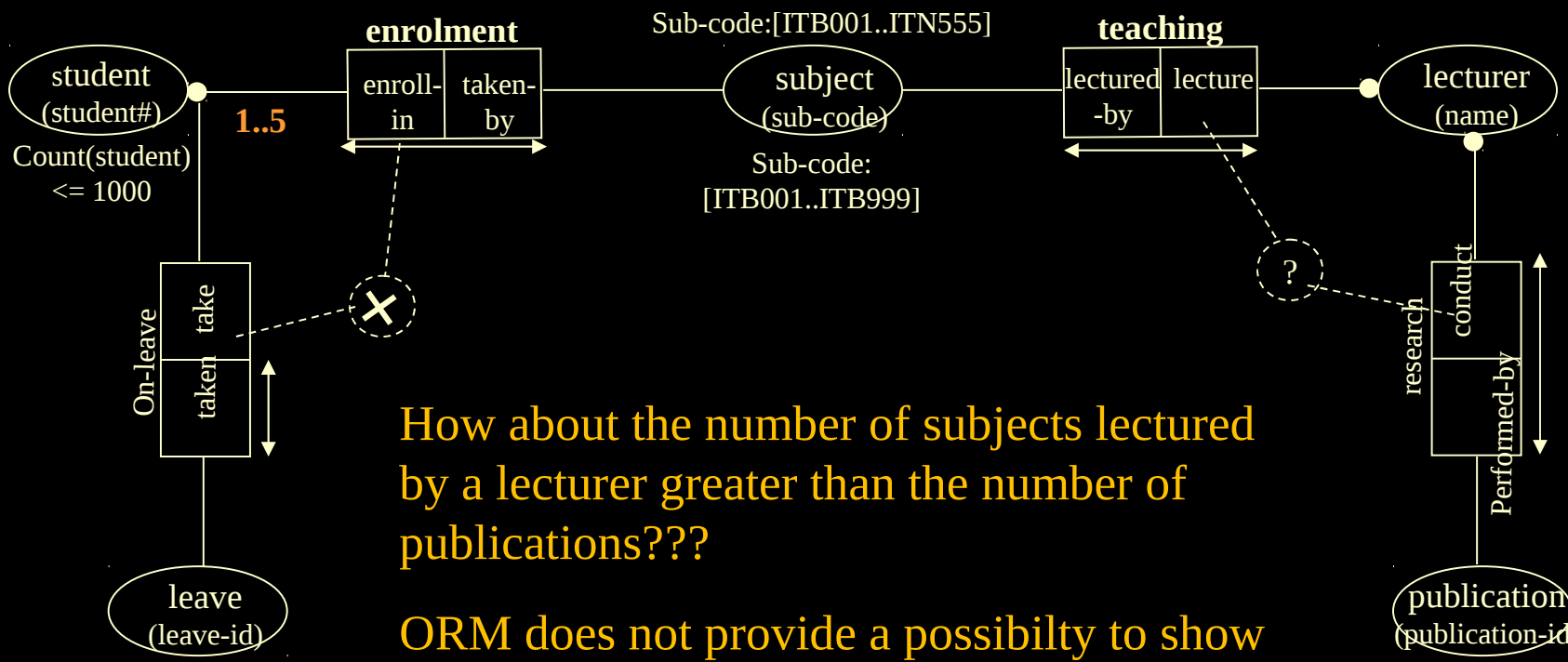
A role may be allowed to hold only for a subset of the values allowed for its entity.
In the following example Money allowes any value permitted by label type real(8,2), whereas Salary can only hold values of Money between 20000 and 80000 and Expenses can be between 500 and 5000. This information can be used in defining triggers in database application



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.

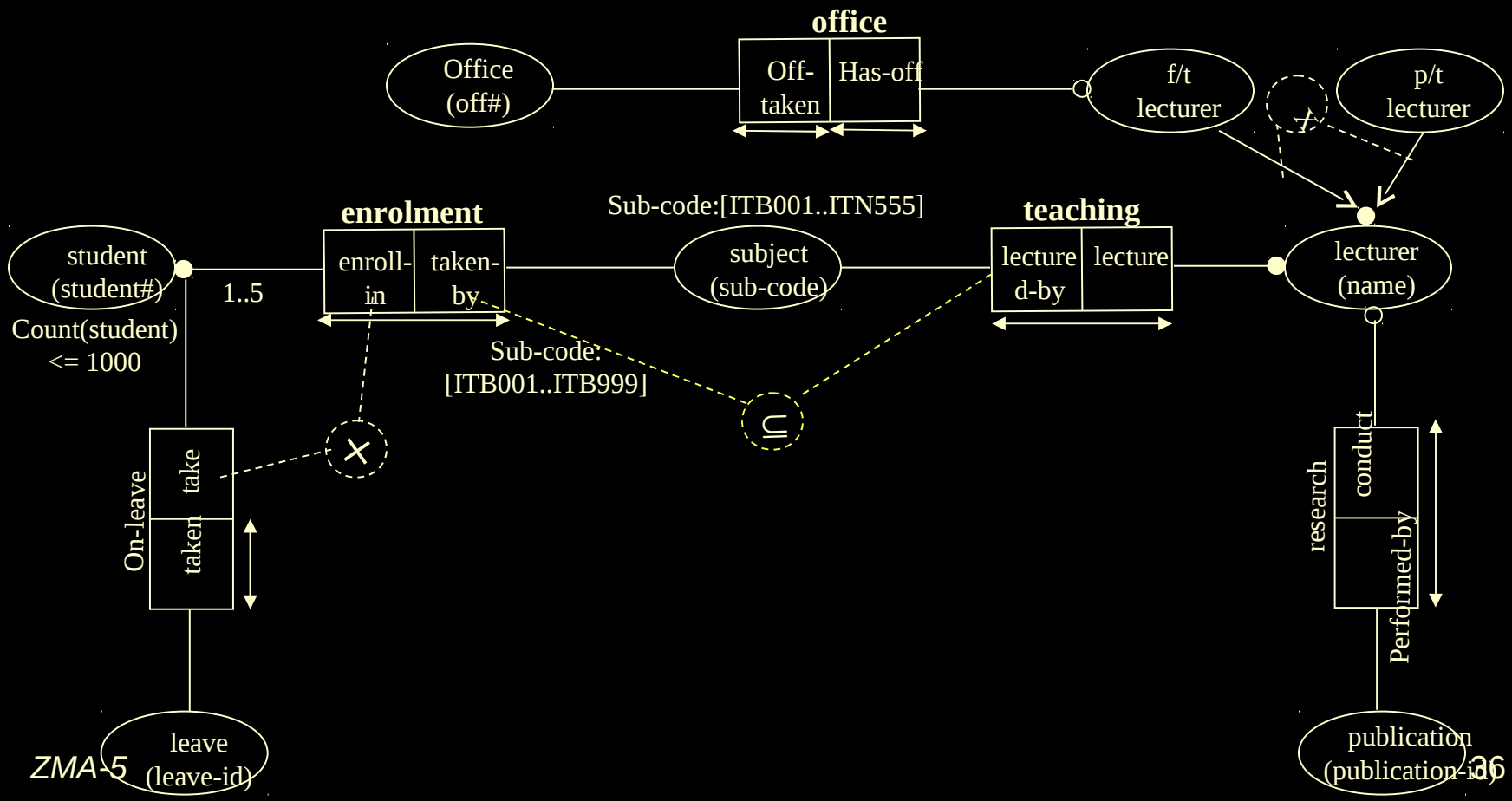


How about the number of subjects lectured by a lecturer greater than the number of publications???

ORM does not provide a possibilty to show it in the above schema

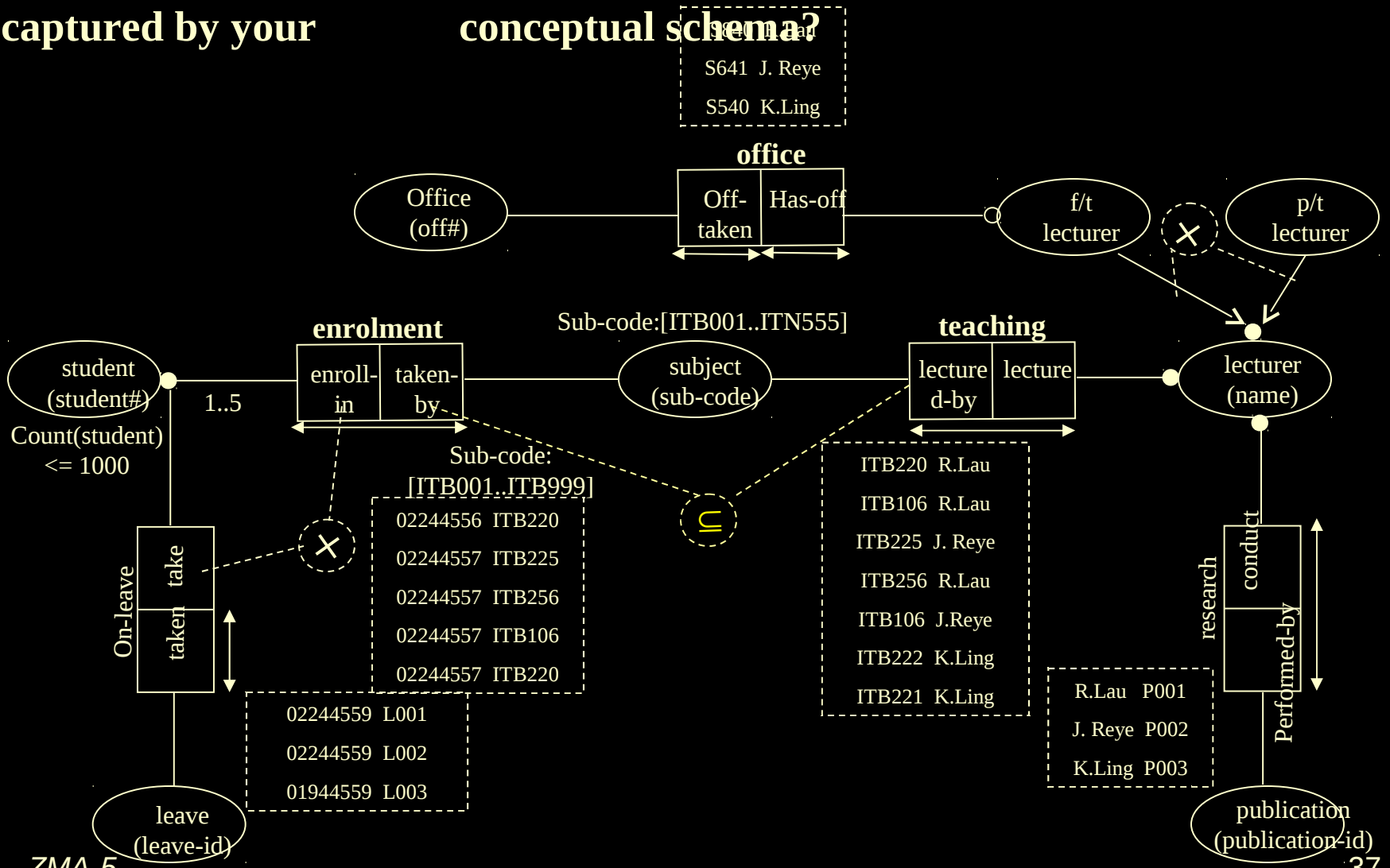
Add other constraints and perform final checks.

- **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)

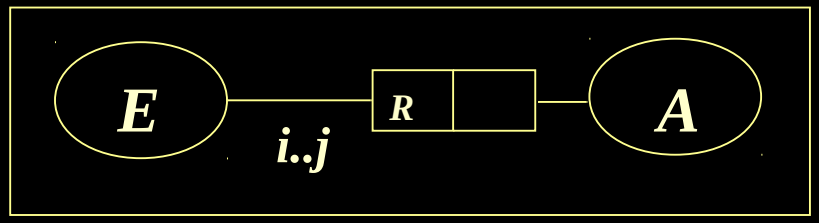
Can the fact type population satisfy all the constraints specified in your conceptual schema? Can all the population (data) in the UoD be captured by your conceptual schema?



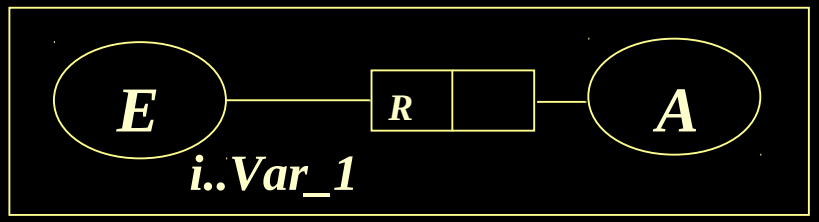
Occurrence Frequencies – more formally

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

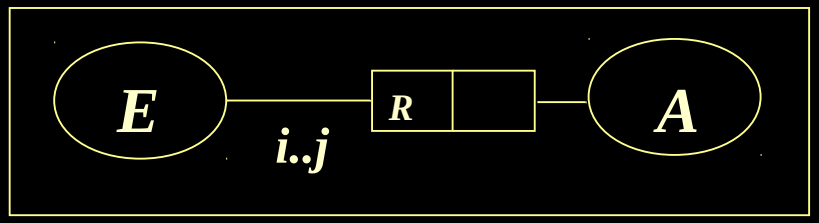
It specifies how often an instance of an entity type, or combination of entity types, may be involved in a role, or combination of roles.



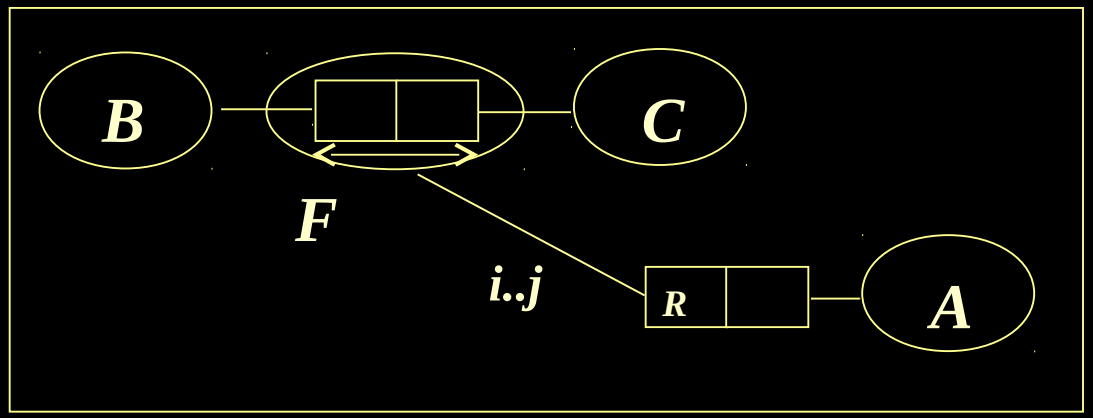
One or both limits in occurrence frequency constraints does not have to be constant – could vary upon some properties of entity types involved in the role affected by that constraint

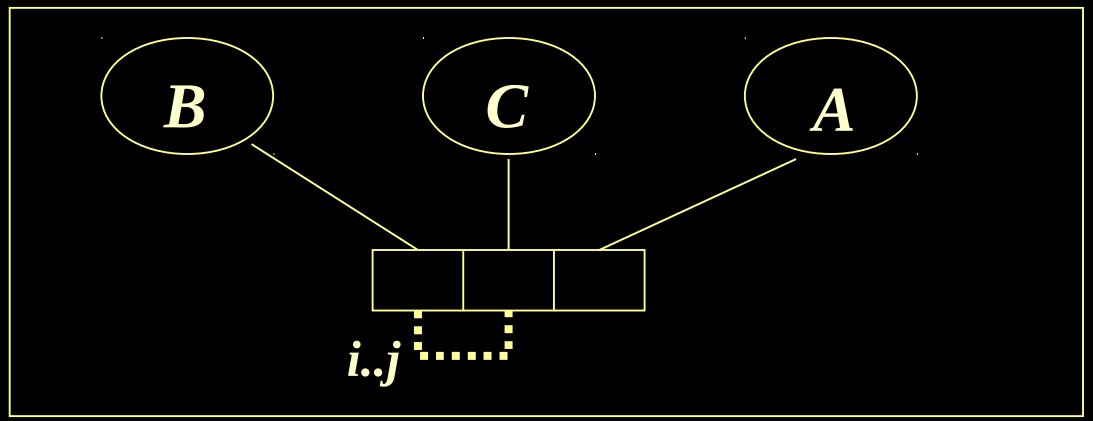


Occurrence Frequencies – for a single role

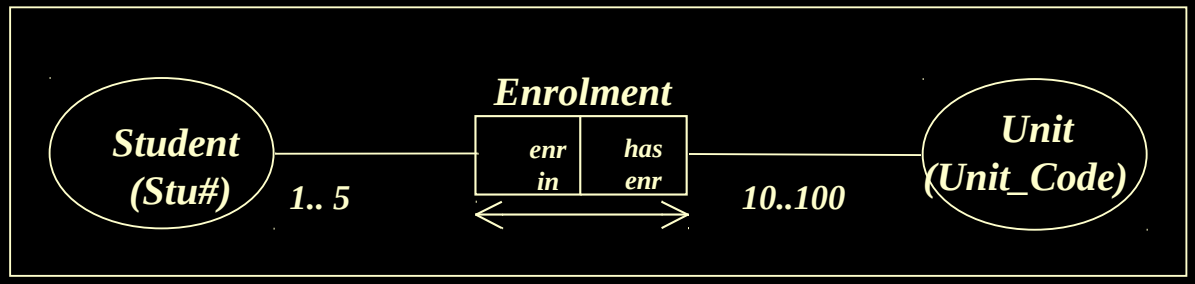


Occurrence Frequencies – for multiple roles





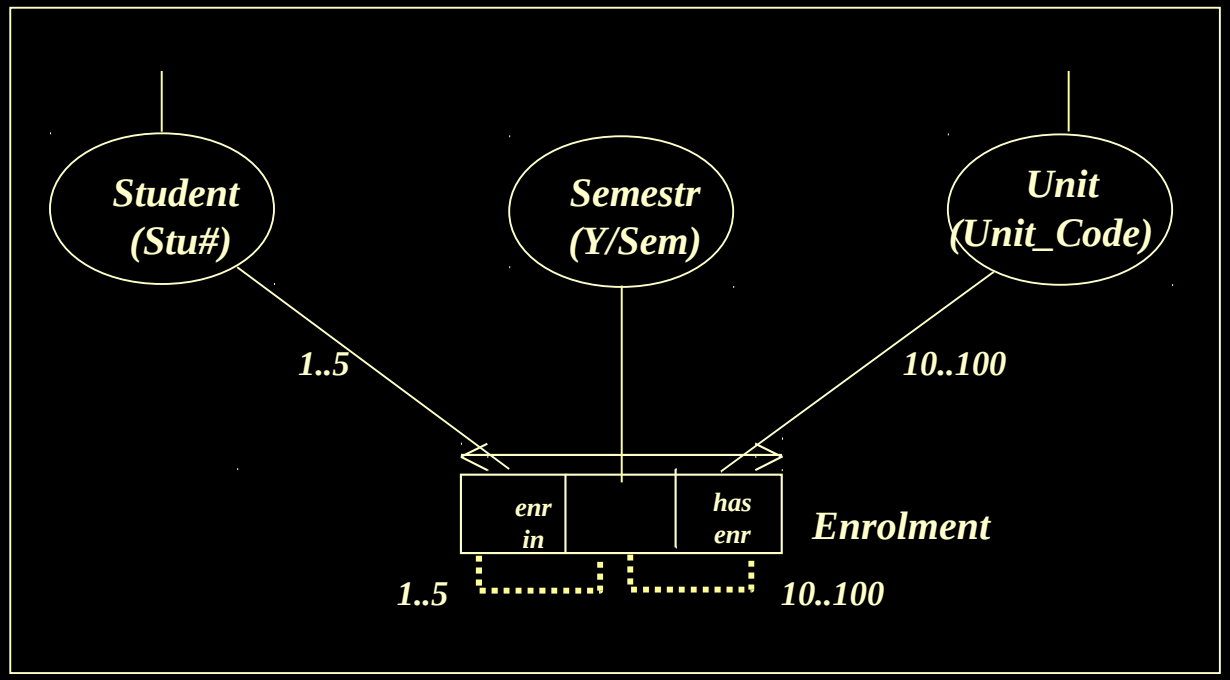
Example



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

Stu#	Unit_code
s1	ITB220
s1	ITB222
s1	ITB223
s3	ITB220
s3	ITB222
s4	ITB220
s4	ITB222
s5	ITB220
s5	ITB222
s6	ITB220
s6	ITB222
s7	ITB220
s7	ITB222
s8	ITB220
s8	ITB222
s9	ITB220
s9	ITB222
s10	ITB220
s10	ITB222
s11	ITB222
s12	ITB222

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 occurence frequency relates to the combination of roles**.**

Population Check

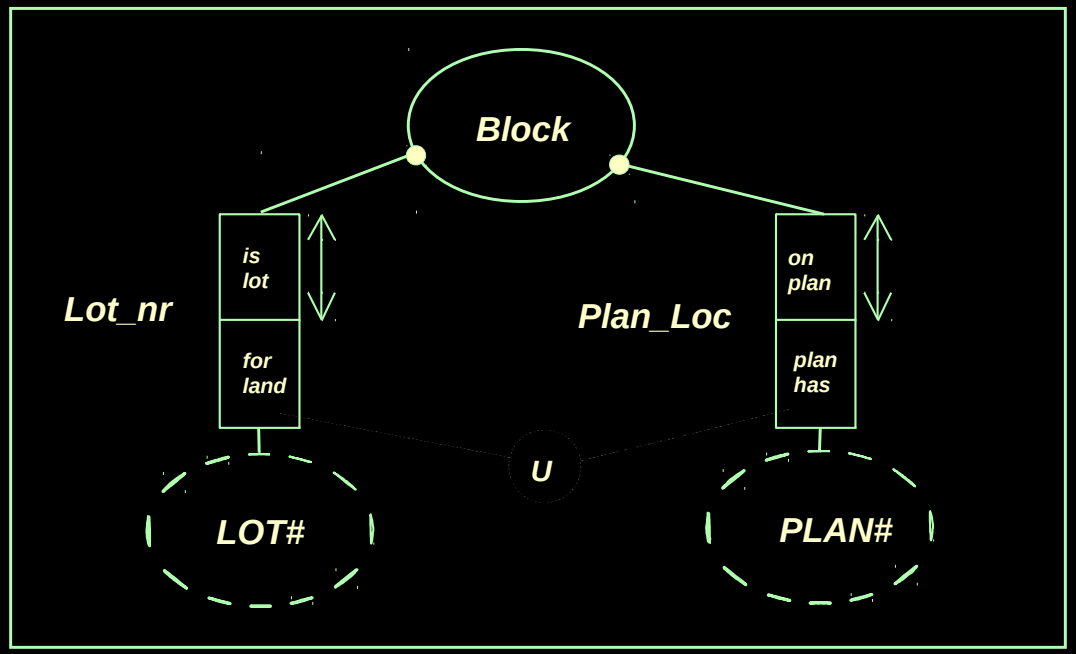
Check if the conceptual schema has too strong constraint anomaly. i.e. is the output report, on which the conceptual schema is based, a permitted population of the conceptual schema?

This can be checked using the natural join operation on the populated fact types.

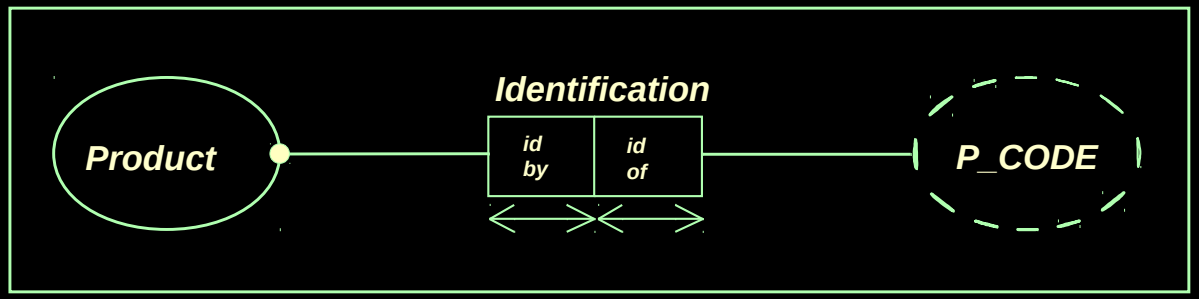
Final checks

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

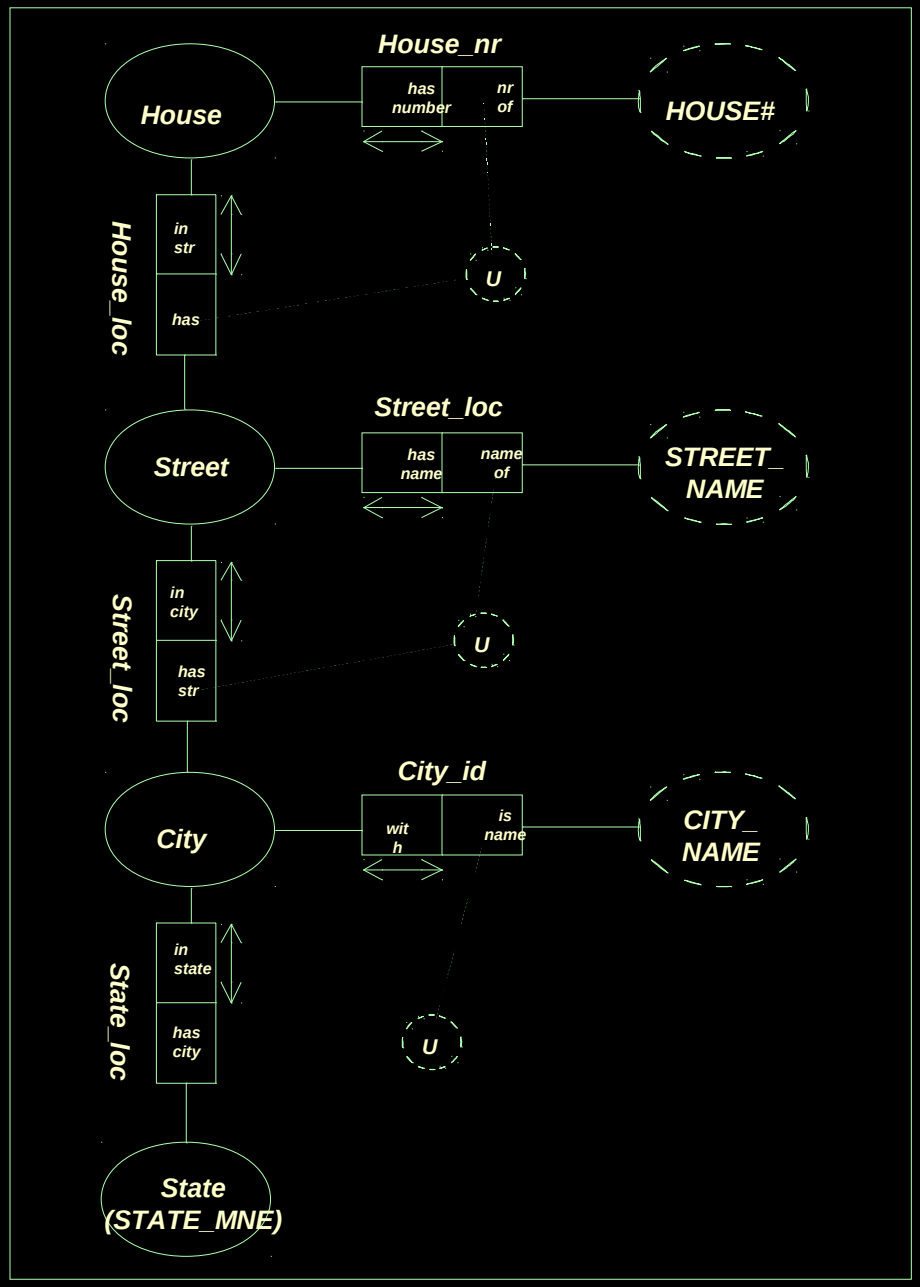


Example



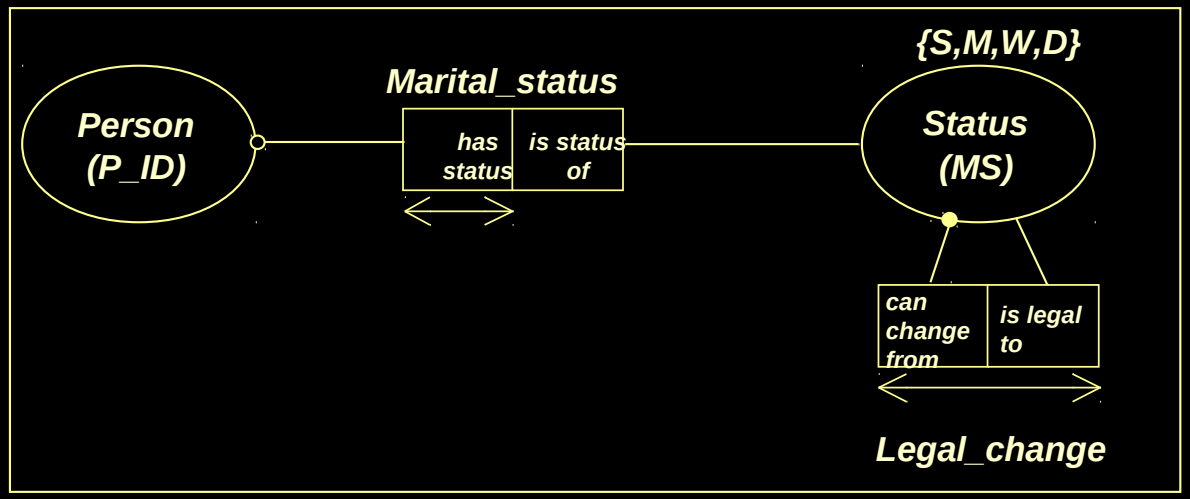
Example

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



S	M
M	W
M	D
D	M
W	M

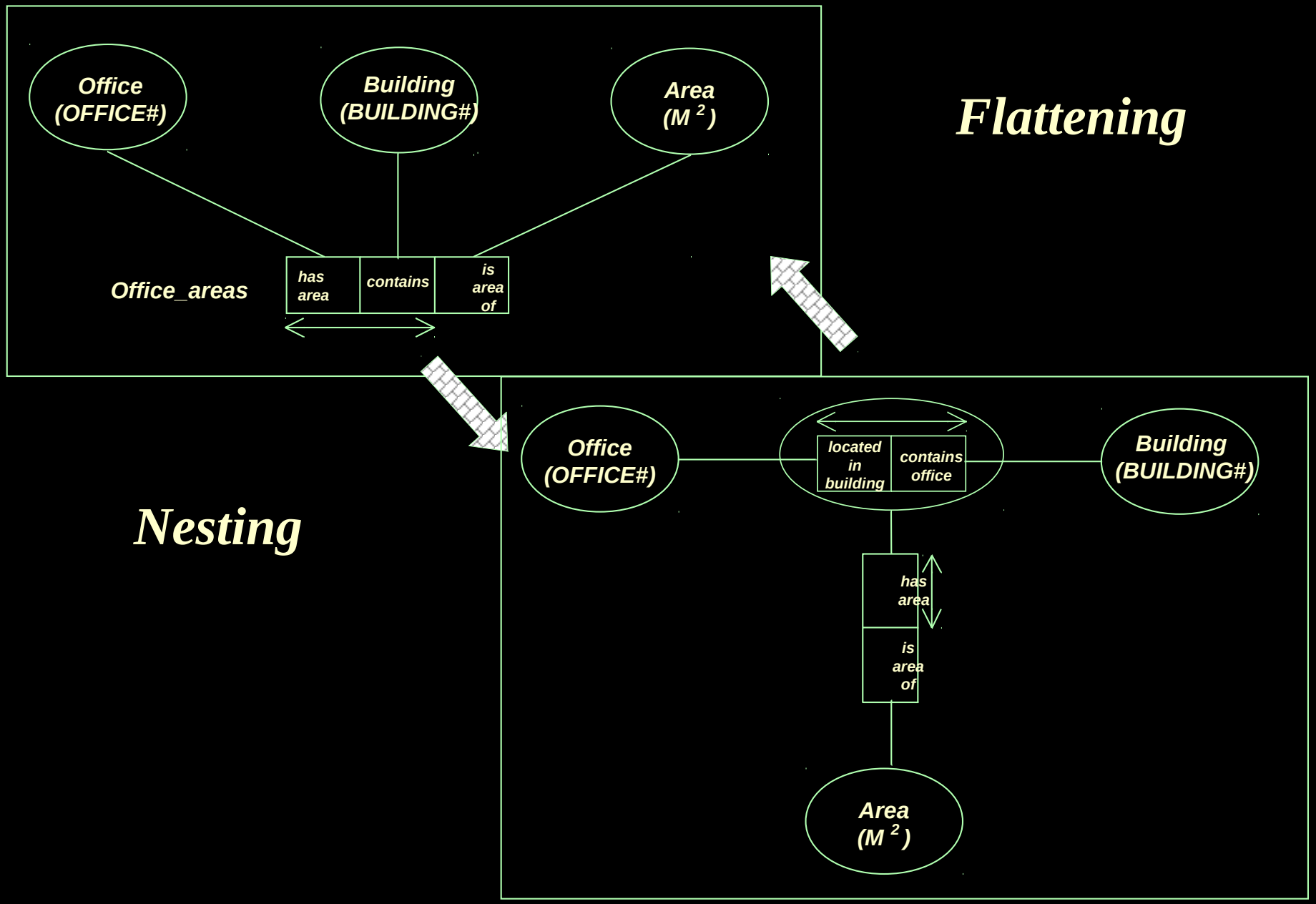
A single person changes to married
A married person changes to divorced or widowed
A divorced person changes to married
A widowed person changes to married.

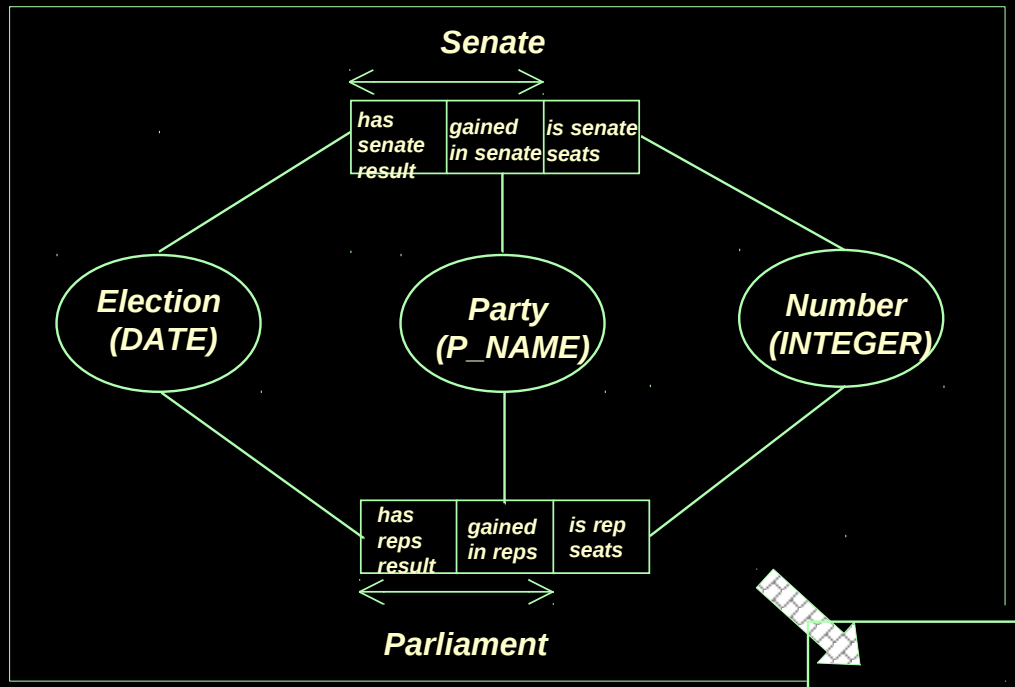
A Legal_change table can be stored in the database and to be used in triggers for checking if updates of Marital status is consistent with legal rules.

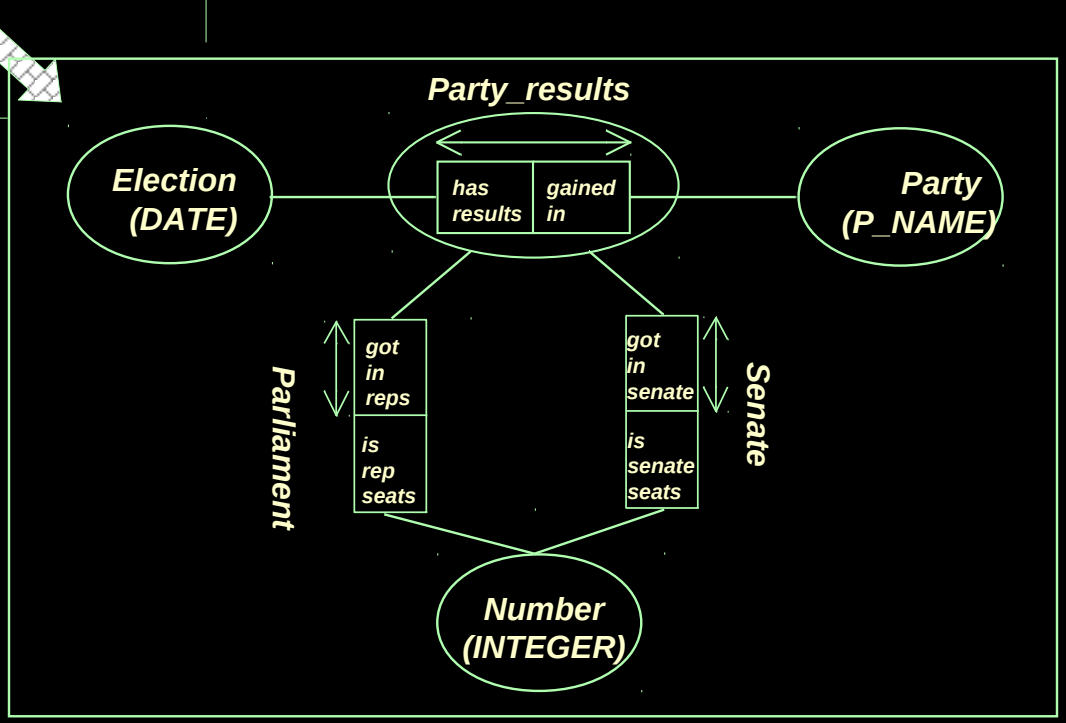
Conceptual Schema Transformation

Different "conceptual views" have the same semantics.

Reductions	*Compositions*
splitting -	*grouping -*
nesting	*flattening*
entity type to role reduction	*role to entity type composition*
fact type to entity type conversion	*entity type to fact type conversion*

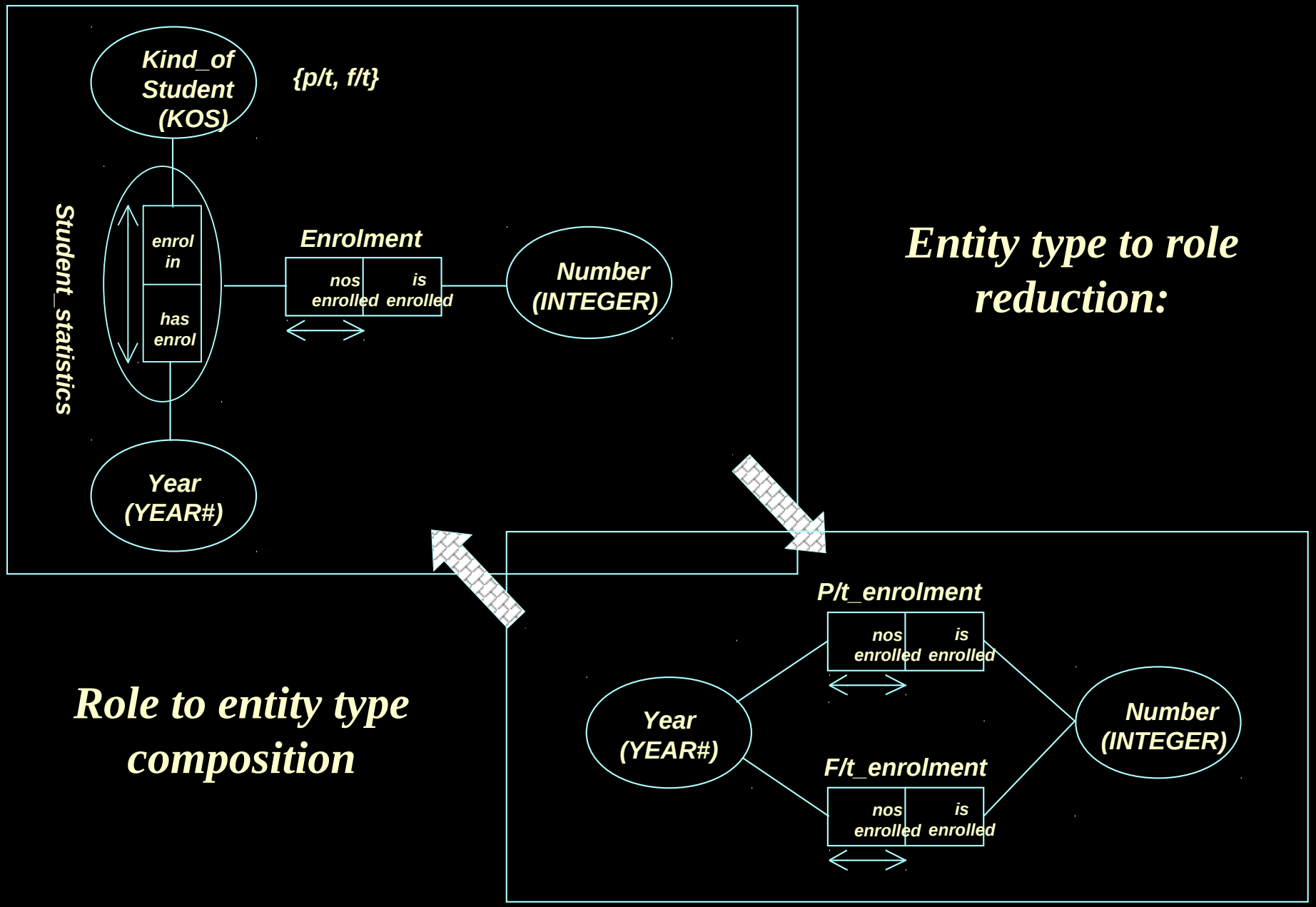

Office
(OFFICE#)
Building
(BUILDING#)
Area
(M 2)
Flattening
Office_areas
has
area
contains
is
area
of
Nesting
Office
(OFFICE#)
located
in
building
contains
office
Building
(BUILDING#)
has
area
is
area
of
Area
(M 2)

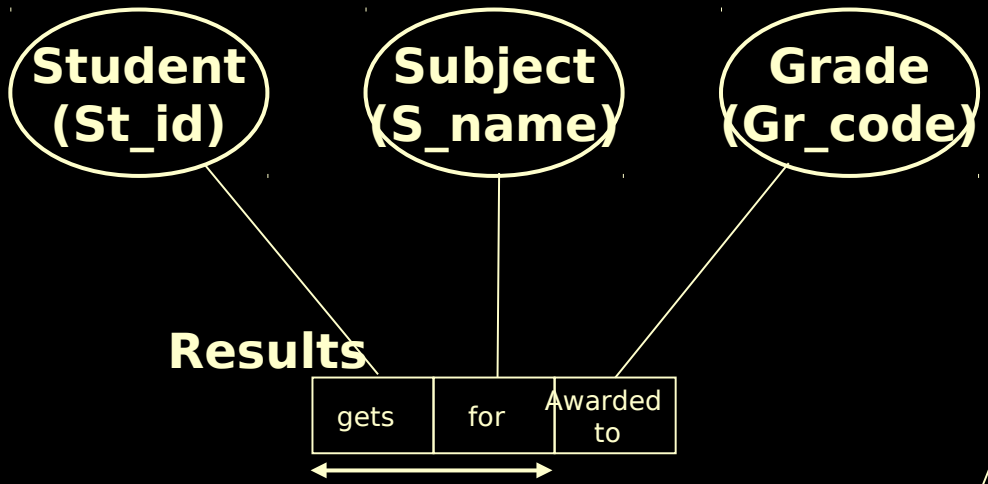


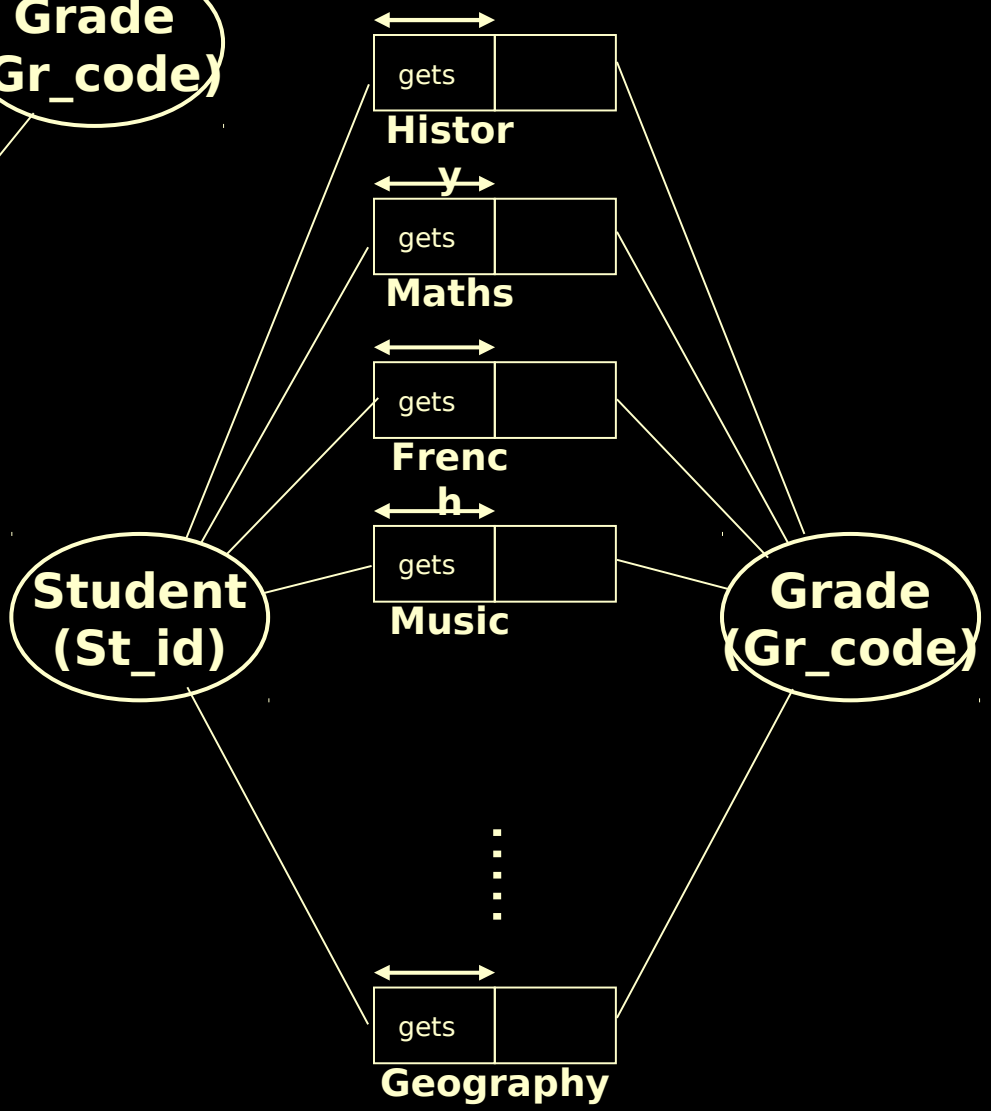


Nesting

Sub-relationship objectified for schema 'optimisation'


Kind_of Student (KOS)
{p/t, f/t}
Student_statistics
enrol in
has enrol
Enrolment
nos enrolled
is enrolled
Number (INTEGER)
Year (YEAR#)
Entity type to role reduction:
Role to entity type composition
P/t_enrolment
nos enrolled
is enrolled
Year (YEAR#)
Number (INTEGER)
F/t_enrolment
nos enrolled
is enrolled

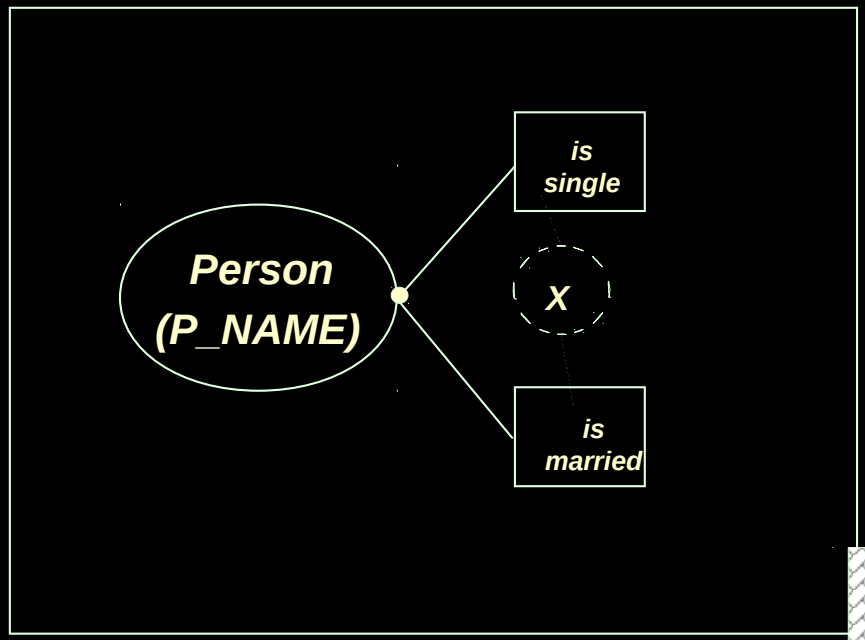




Entity type to role reduction

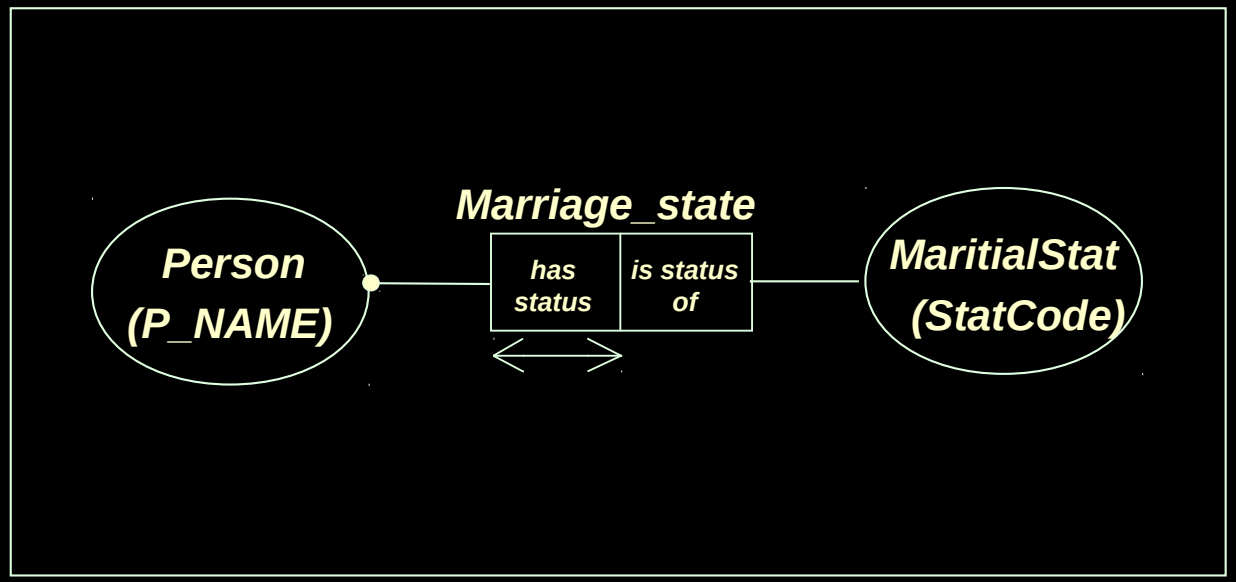
Different schemas for the same UoD.

Resulting in different structure of data storage.



Fact type to entity type conversion
Jumping transition

Preferred representation:



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