Conceptual Data Modeling Concepts & Principles

(Chapter 1&2)

Mustafa Jarrar

Birzeit University

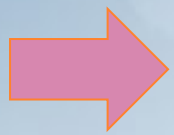
mjarrar@birzeit.edu
www.jarrar.info

Watch this lecture and download the slides

Course Page: http://www.jarrar.info/courses/KnowledgeEngineering/

More Online Courses at: http://www.jarrar.info/courses/

Outline



Information Modeling

- Conceptual Modeling Approaches
- Introduction to Object Role Modeling (ORM)
- Information Levels

Information Modeling –The need for good design

Do you like the design of this table?

Movie	*Year*	*Director*	*Stars*
Awakenings	1991	Penny Marshall	Robert De Niro Robin Williams
Backdraft	1991	Ron Howard	William Baldwin Robert De Niro Kurt Russell
Cosmology	1994	Terry Harding	
Dances with wolves	1990	Kevin Kostner	Kevin Kostner Mary McDonnell

- This table is an output report. It provides one way to view the data.
- Different movies may have the same title.
- Movie numbers are used to provide a simple identifier.
- Each cell (row--column slot) may contain many values.

➢How can we design tables to store such facts?

Information Modeling –The need for good design

A badly-designed table, why?

Movie

MovieName	Release Year	Director	Star
Awakenings	1991	Penny Marshall	Robert De Niro
Awakenings	1991	Penny Marshall	Robin Williams
Backdraft	1991	Ron Howard	William Baldwin
Backdraft	1991	Ron Howard	Robert De Niro
Backdraft	1991	Ron Howard	Kurt Russell
Cosmology	1994	Terry Harding	
Dances with wolves	1990	Kevin Kostner	Kevin Kostner
Dances with wolves	1990	Kevin Kostner	Mary McDonnell

Information Modeling –The need for good design

Do you like the design of this table?

Movie

MovieName	Release Year	Director
Awakenings	1991	Penny Marshall
Backdraft	1991	Ron Howard
Cosmology	1994	Terry Harding
Dances with wolves	1990	Kevin Kostner

Movie Stars

MovieName	Star
Awakenings	Robert De Niro
Awakenings	Robin Williams
Backdraft	William Baldwin
Backdraft	Robert De Niro
Backdraft	Kurt Russell
Cosmology	
Dances with wolves	Kevin Kostner
Dances with wolves	Mary McDonnell

Information Modeling –The need for good design

Do you like the design of this table?

Movie

MovieName	Release Year	Director
Awakenings	1991	Penny Marshall
Backdraf		
Cosmolo		
Dances w		

Movie Stars

MovieN	
Awakeni	
Awakeni	
Backdra	
Backdra	
Backdraft	Kurt Russell
Cosmology	
Dances with wolves	Kevin Kostner
Dances with wolves	Mary McDonnell

A relational database representation

Movie (movieName, releaseYr, director)

Starred (movieName, star)

Information Modeling –The need for good design

Do you design of this table?

Movie

Mo	
Aw	
Ba	
Co	
Da	

Movie Stars

M	
Awakenings	Robin Williams
Backdraft	William Baldwin
Backdraft	Robert De Niro
Backdraft	Kurt Russell
Cosmology	
Dances with wolves	Kevin Kostner
Dances with wolves	Mary McDonnell

- Information Modeling is both a science and an art.
- When supported by a good modeling approach, this design process is a stimulating and intellectually satisfying activity, with tangible benefits gained from the quality of the database applications produced.

Information Modeling –The need for good design

- Why a good design is important?
  - Consistency
  - Efficiency
- What makes a good design good?
  - Correct
  - Complete
  - Efficient
- What skills you should have to be a good information modeler?
- What approaches exist to help you reach good models?

Information Modeling

- The application area being modeled is called the universe of discourse (UoD).
- Building a good model requires a good understanding of the world we are modeling.
- The main challenge is to describe the UoD clearly and precisely.
- A person responsible for modeling the UoD is called a *modeler.*
- we should consult with others who, at least collectively, understand the application domain—these people are called *domain experts, subject* matter experts, or UoD experts.
- For implementation, it is important to represent information at the conceptual level -in concepts that people (molders and domain experts) find easy to work with.
- This added flexibility also makes it easier to implement the same conceptual model in different ways, DB schema, XML schema, etc.

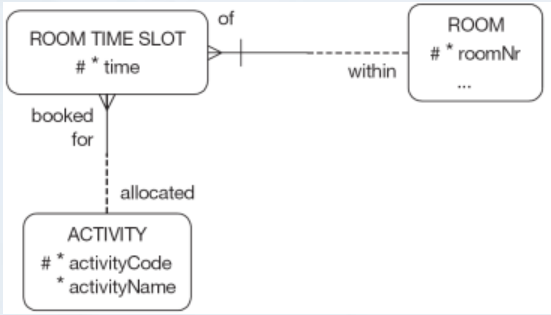
Outline

- Information Modeling
- Modeling Approaches
- Introduction to Object Role Modeling (ORM)
- Information Levels

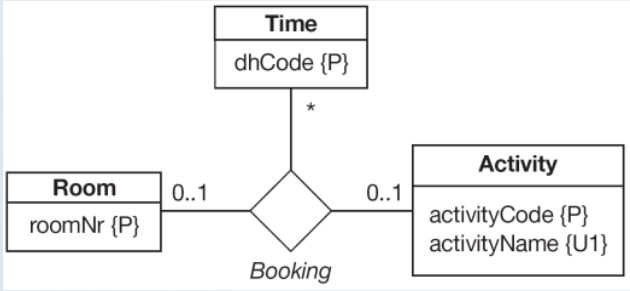
Modeling Approaches

The main information modeling approaches are:

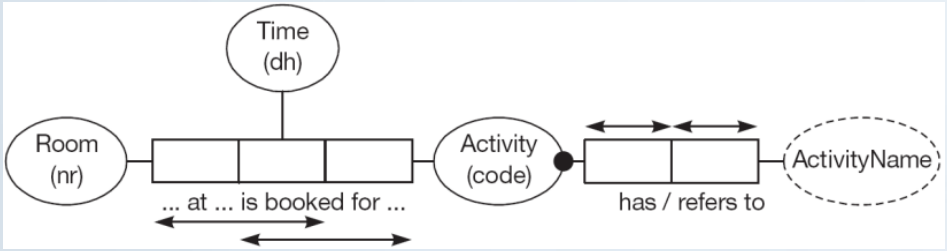
- Entity-Relationship modeling (ER)



- Object-oriented modeling (UML)



- Fact-oriented modeling (ORM)

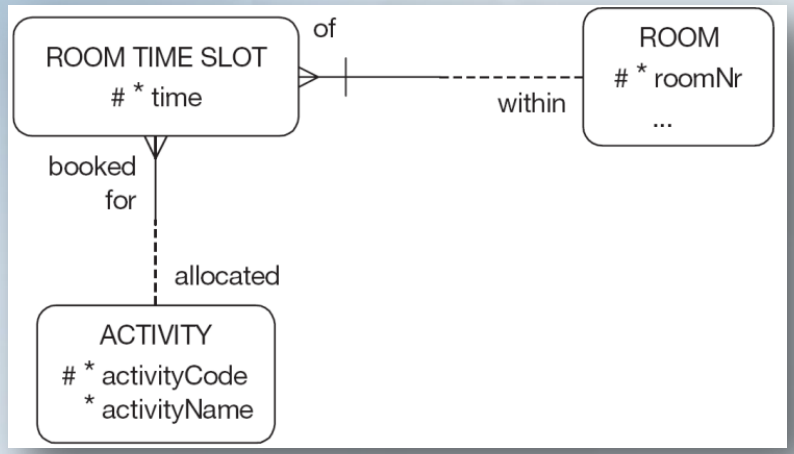


Modeling Approaches

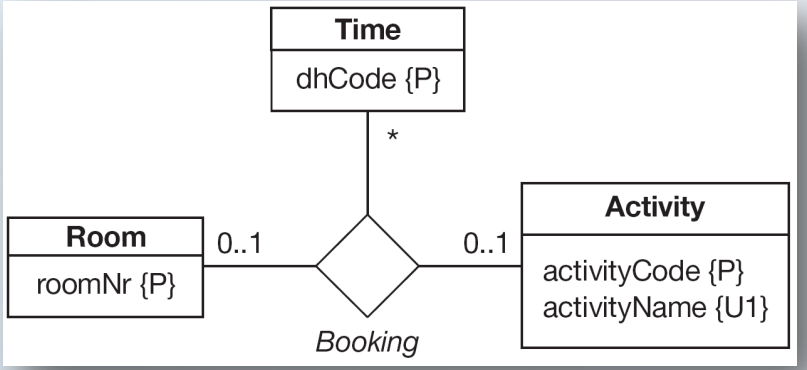
Given simple data for room scheduling:

Room	*Time*	*ActivityCode*	*ActivityName*
20	Mon 9 a.m.	VMC	VisioModeler class
20	Tue 2 p.m.	VMC	VisioModeler class
33	Mon 9 a.m.	AQD	ActiveQuery demo
33	Fri 5 p.m.	SP	Staff party
...	...	...	...

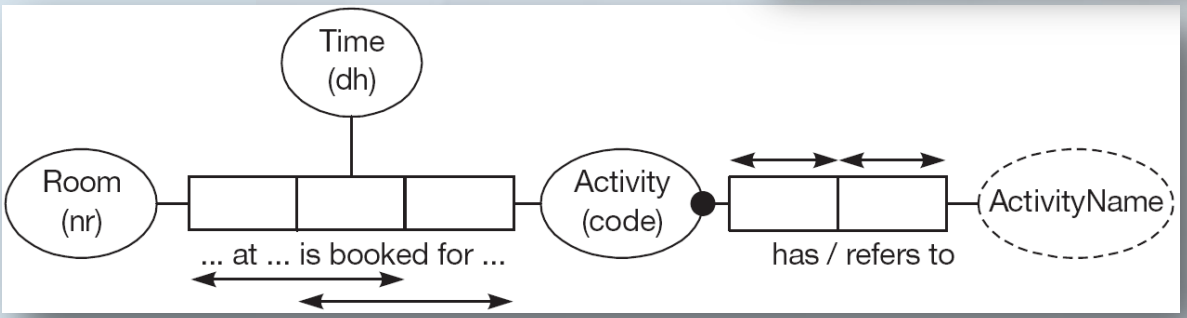
ER-model



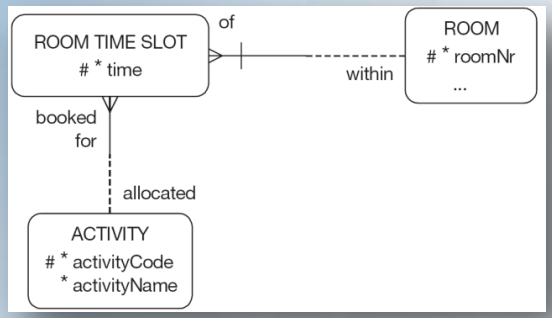
UML-model



ORM-model



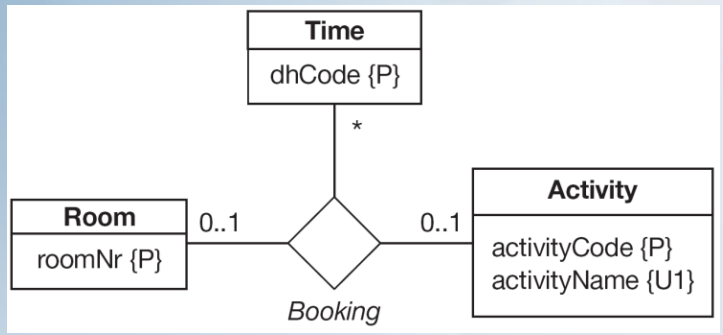
Entity-Relationship Modeling (ER)



Room	*Time*	*ActivityCode*	*ActivityName*
20	Mon 9 a.m.	VMC	VisioModeler class
20	Tue 2 p.m.	VMC	VisioModeler class
33	Mon 9 a.m.	AQD	ActiveQuery demo
33	Fri 5 p.m.	SP	Staff party
...	...	...	...

- Introduced by Peter Chen in 1976, widely used approach for DB modeling.
- Pictures the world in terms of entities that have attributes and participate in relationships.
- Many different versions of ER (no standard ER notation). Different versions of ER may support different concepts and may use different symbols for the same concept.
- Relationships are depicted as named lines connecting entity types. Only binary relationships are allowed, and each half of the relationship is shown either as a solid line (mandatory) or broken line (optional). A fork or "crow's foot" at one end of a relationship indicates that many instances of the entity type at that end may be associated (via that relationship) with the same entity instance at the other end of the relationship. The lack of a crow's foot indicates that at most one entity instance at that end is associated with any given entity instance at the other end.

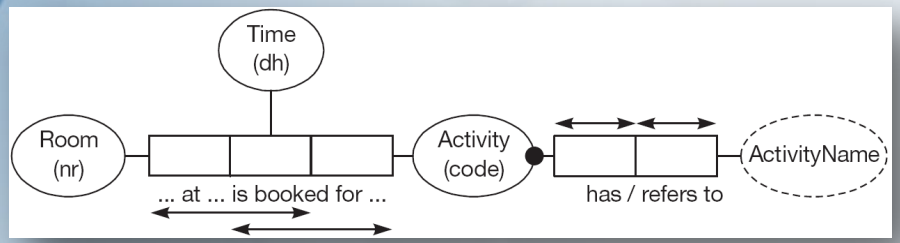
Object-oriented Modeling (UML)



Room	*Time*	*ActivityCode*	*ActivityName*
20	Mon 9 a.m.	VMC	VisioModeler class
20	Tue 2 p.m.	VMC	VisioModeler class
33	Mon 9 a.m.	AQD	ActiveQuery demo
33	Fri 5 p.m.	SP	Staff party
...	...	...	...

- UML class diagram are used to specify static data structures (OMG Standard).
- Encapsulates both data and behavior within objects.
- Pictures the world in terms of classes that have attributes and participate in associations. Ternary associations are allowed, see the diagram.
- UML allow constraints in braces or notes in whatever language you wish.
- Form example, {P} can be added to denote primary uniqueness and {U1} for an alternate uniqueness—these symbols are not standard and hence not portable. The uniqueness constraints on the ternary are captured by the two 0..1 (at most one) multiplicity constraints. The "*" means "0 or more". Attributes are mandatory by default.

Fact-oriented Modeling (ORM)



Room	*Time*	*ActivityCode*	*ActivityName*
20	Mon 9 a.m.	VMC	VisioModeler class
20	Tue 2 p.m.	VMC	VisioModeler class
33	Mon 9 a.m.	AQD	ActiveQuery demo
33	Fri 5 p.m.	SP	Staff party
...	...	...	...

- Introduced by Sjir Nijssen early 1970s, was called NIAM.
- Revised by Terry Halpin (late 1980s), and called:

Object-Role Modeling (ORM)

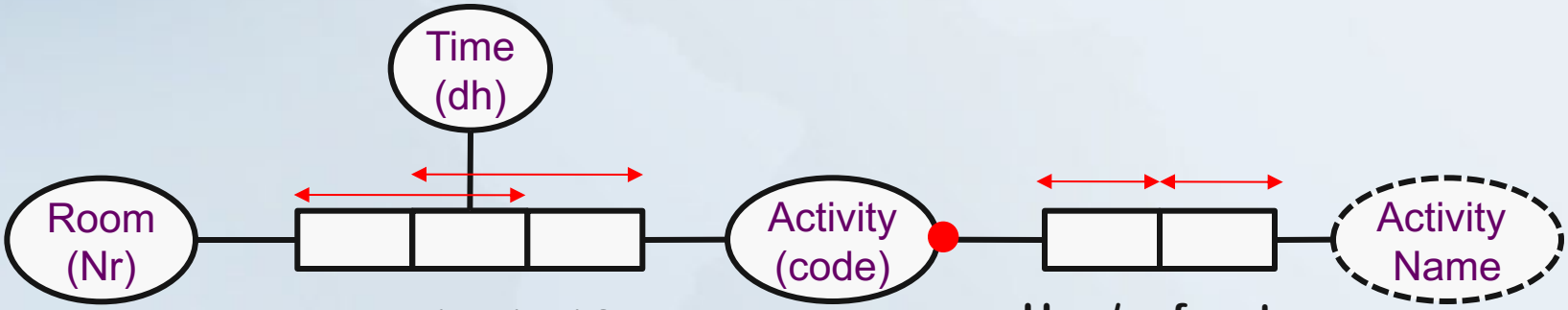
- It views the world as object-types playing roles.
- Object-types are ellipses (no attributes), and relations consists of roles.
- Not only n-ary relations are supported, but ORM supports also more than 15 types of constrains graphically.
- ORM allows verbalization of diagrams.
- More conceptual than UML and ER.
- ORM is a modeling approach, not only a modeling language.

Outline

- Information Modeling
- Modeling Approaches
- Introduction to Object Role Modeling (ORM)
- Information Levels

Object-Role Modeling (ORM)

Room	Time	Activity Code	Activity Name
20	Mon 9am	VMC	VisioModeler Class
20	Tue 2pm	VMC	VisioModeler Calss
33	Mon 9am	AQD	ActiveQuery Demo
33	Fri 5 pm	SP	Staff party
…	…	…	…



…at … is booked for…

20	Mon 9am	VMC
20	Tue 2pm	VMC
33	Mon 9am	AQD
33	Fri 5pm	SP
…	…	…

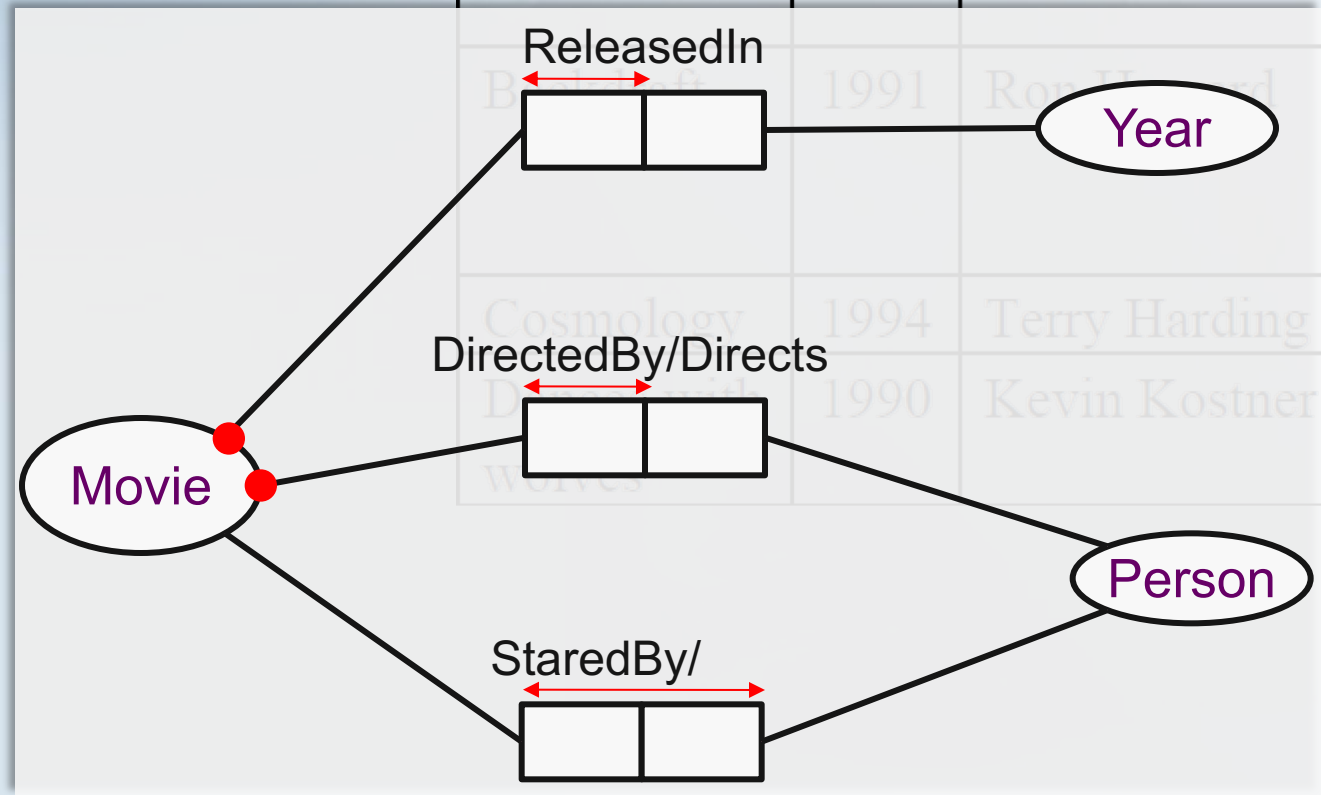
Has / refers to

AQD	ActiveQuery Demo
SP	Staff party
VMC	VisioModeler Calss
VMC	VisioModeler Class
Y2K	Year 200 seminar
…	…

Object-Role Modeling (ORM)

Representing information graphically

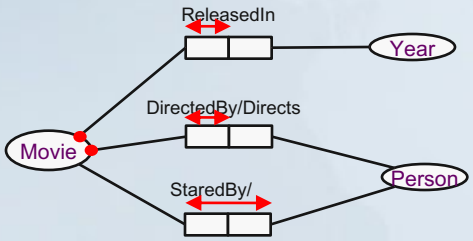
Movie	*Year*	*Director*	*Stars*
Awakenings	1991	Penny Marshall	Robert De Niro Robin Williams
			William Baldwin Robert De Niro Kurt Russell
			Kevin Kostner Mary McDonnell

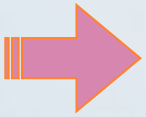


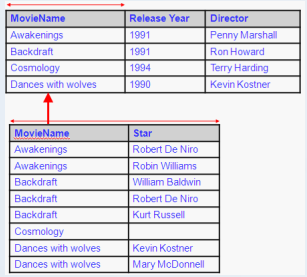
Object-Role Modeling (ORM)

- ORM is conceptual modeling language.
- ORM has an expressive graphical notation.
- ORM is designed for modeling DB schemes at the conceptual level.

  You build an ORM schema and then click to automatically generate a database.







MovieName	Release Year	Director
Awakenings	1991	Penny Marshall
Backdraft	1991	Ron Howard
Cosmology	1994	Terry Harding
Dances with wolves	1990	Kevin Kostner

MovieName	Star
Awakenings	Robert De Niro
Awakenings	Robin Williams
Backdraft	William Baldwin
Backdraft	Robert De Niro
Backdraft	Kurt Russell
Cosmology	
Dances with wolves	Kevin Kostner
Dances with wolves	Mary McDonnell

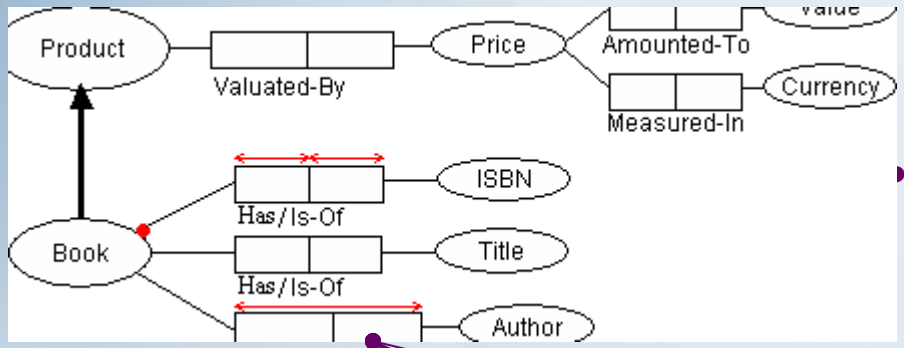
Conceptual Level

Logical Level

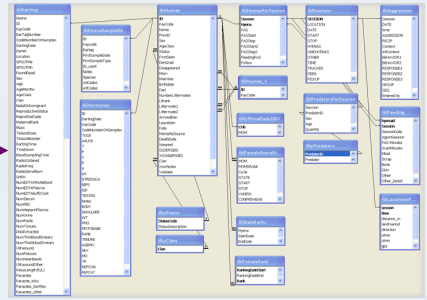
➔Our goal in this course is to use ORM as general Conceptual Modeling language, rather than as database modeling language.

➔ ORM can be used for modeling business rules, ontology, XML schemes, and others.

ORM Usage Scenarios



Originally

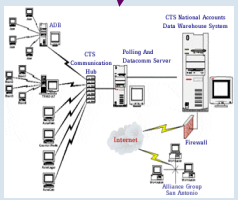


Database

Later

Ontology

XML Schema

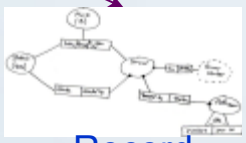


Warehouse

Business Rules

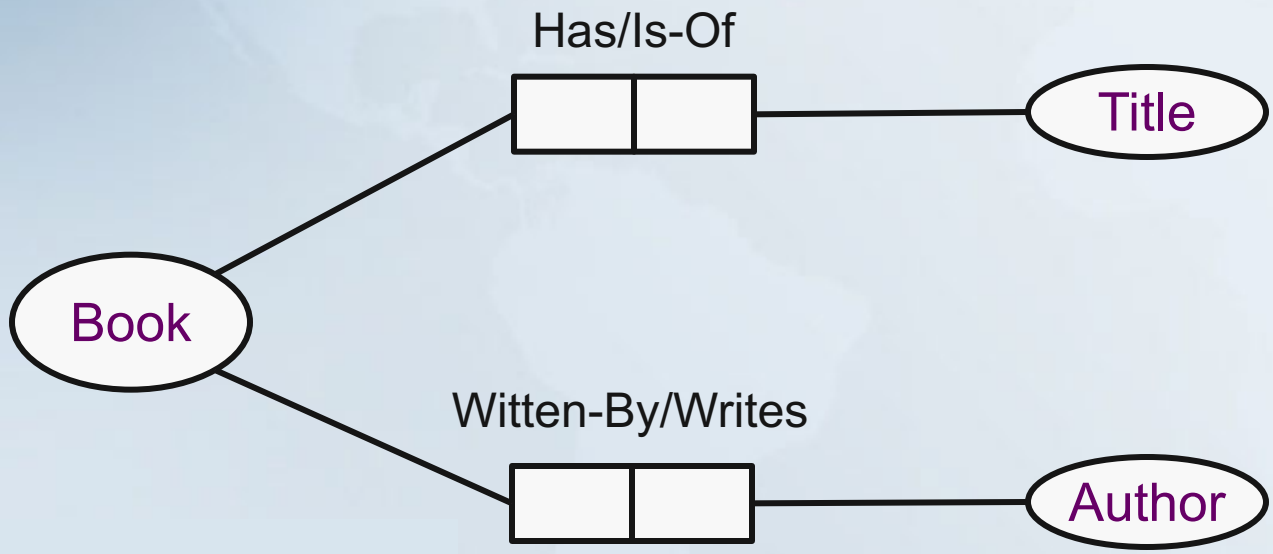
Web (x)Forms

Requirements Engineering

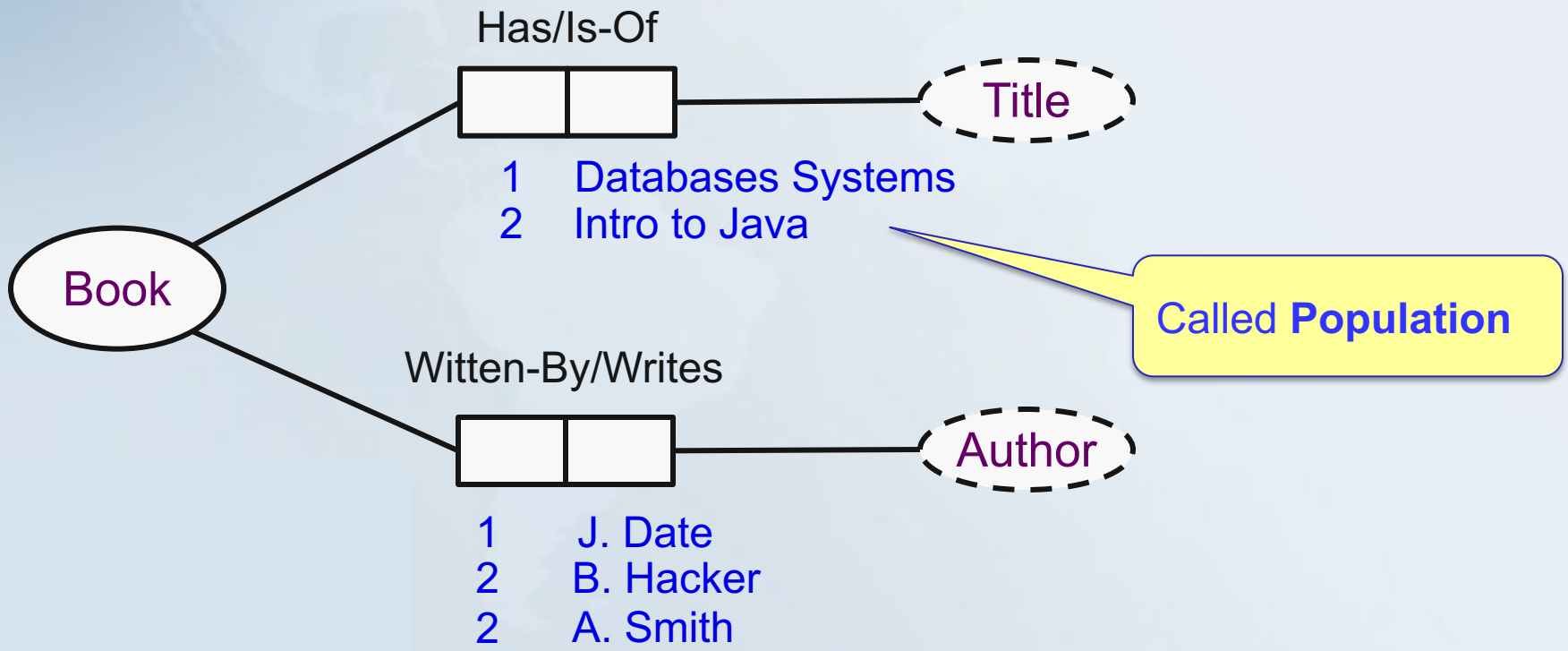


Record my recipes !

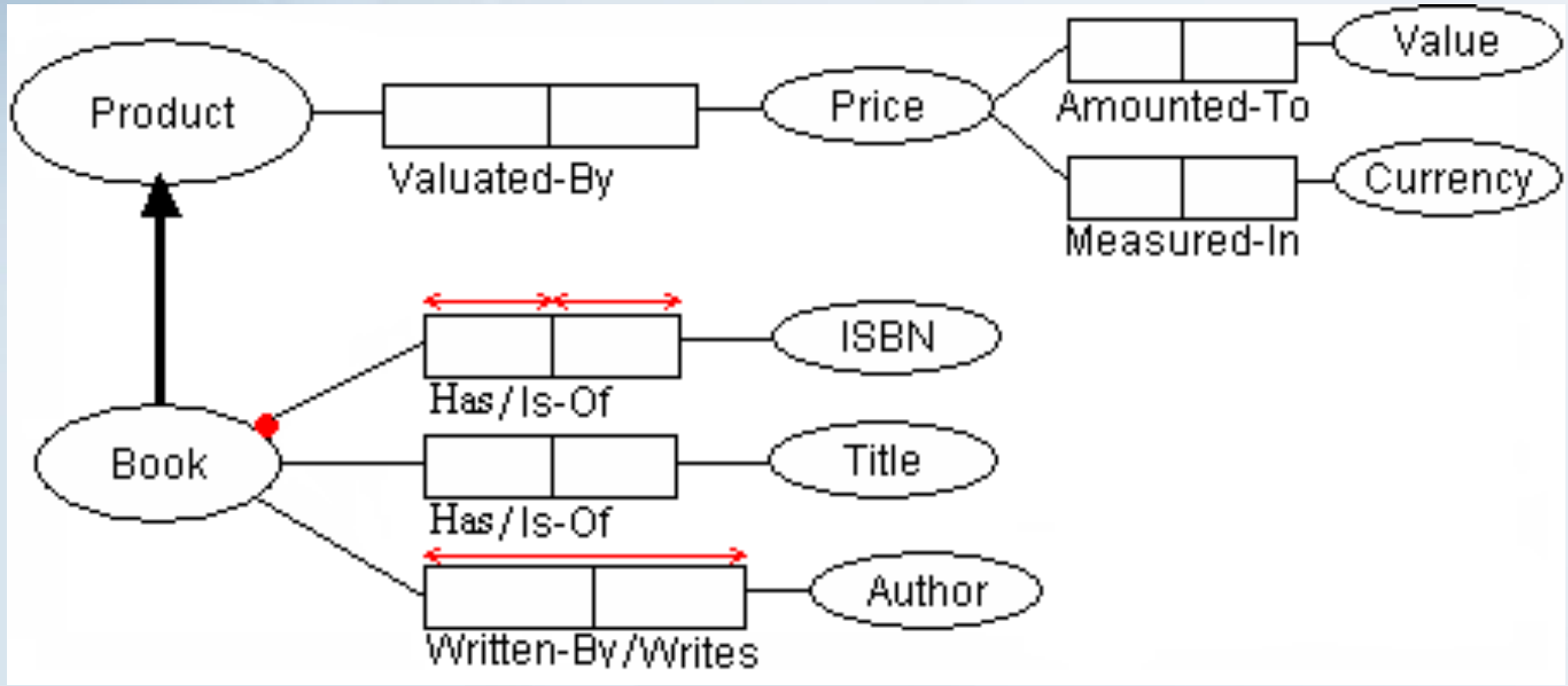
Object-Role Modeling (ORM): Other Examples



Object-Role Modeling (ORM): Other Examples



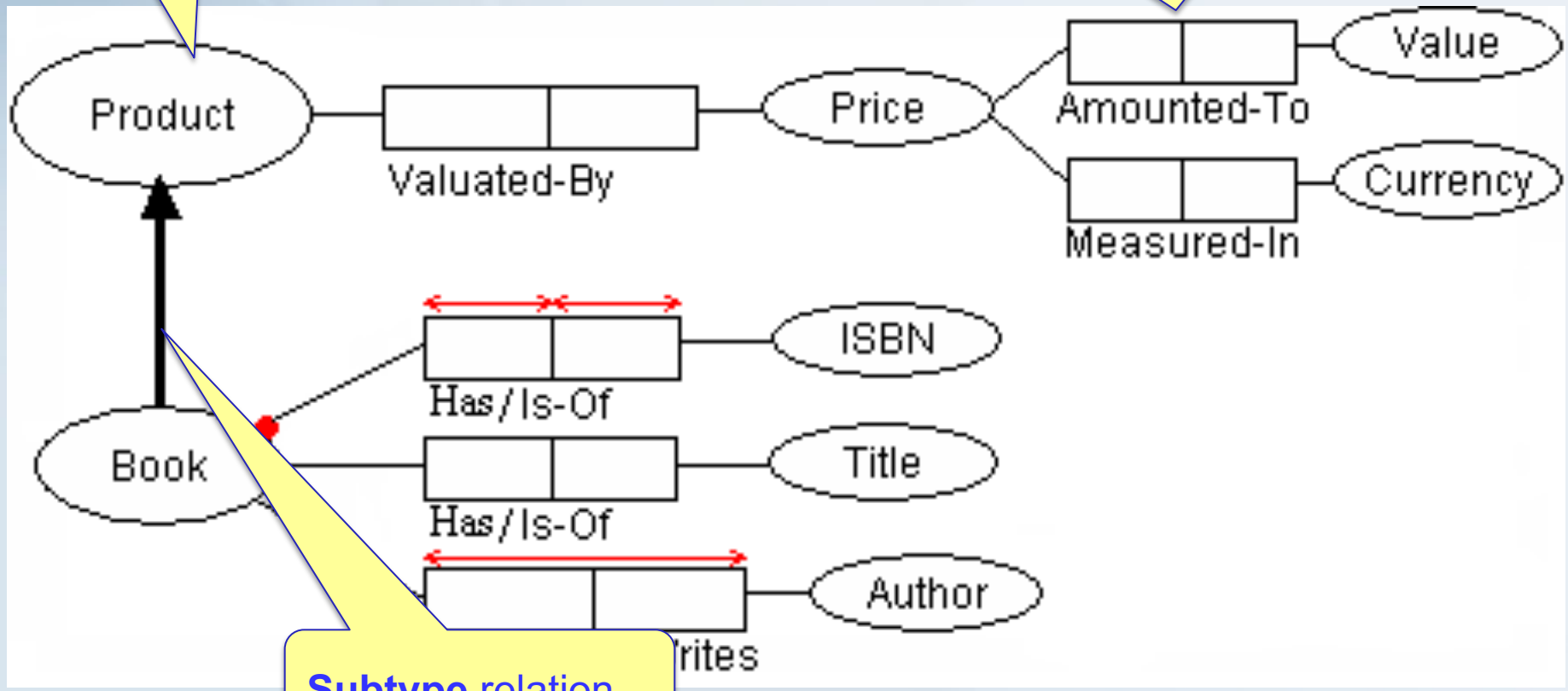
Object-Role Modeling (ORM): Other Examples



Object-Role Modeling (ORM) constructs

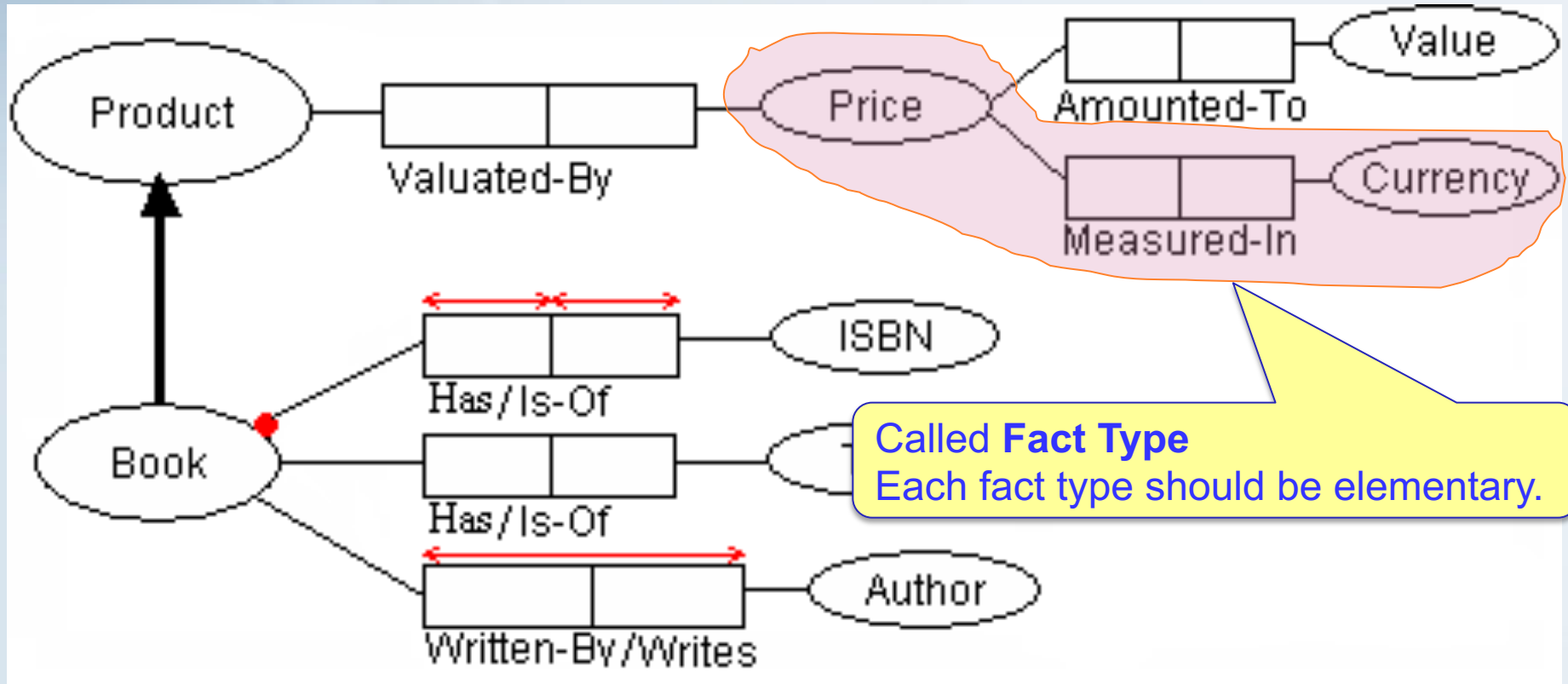
Called **Object Type**
(or Concept, or Class)

Relation
Each part is called a **Role**



Subtype relation

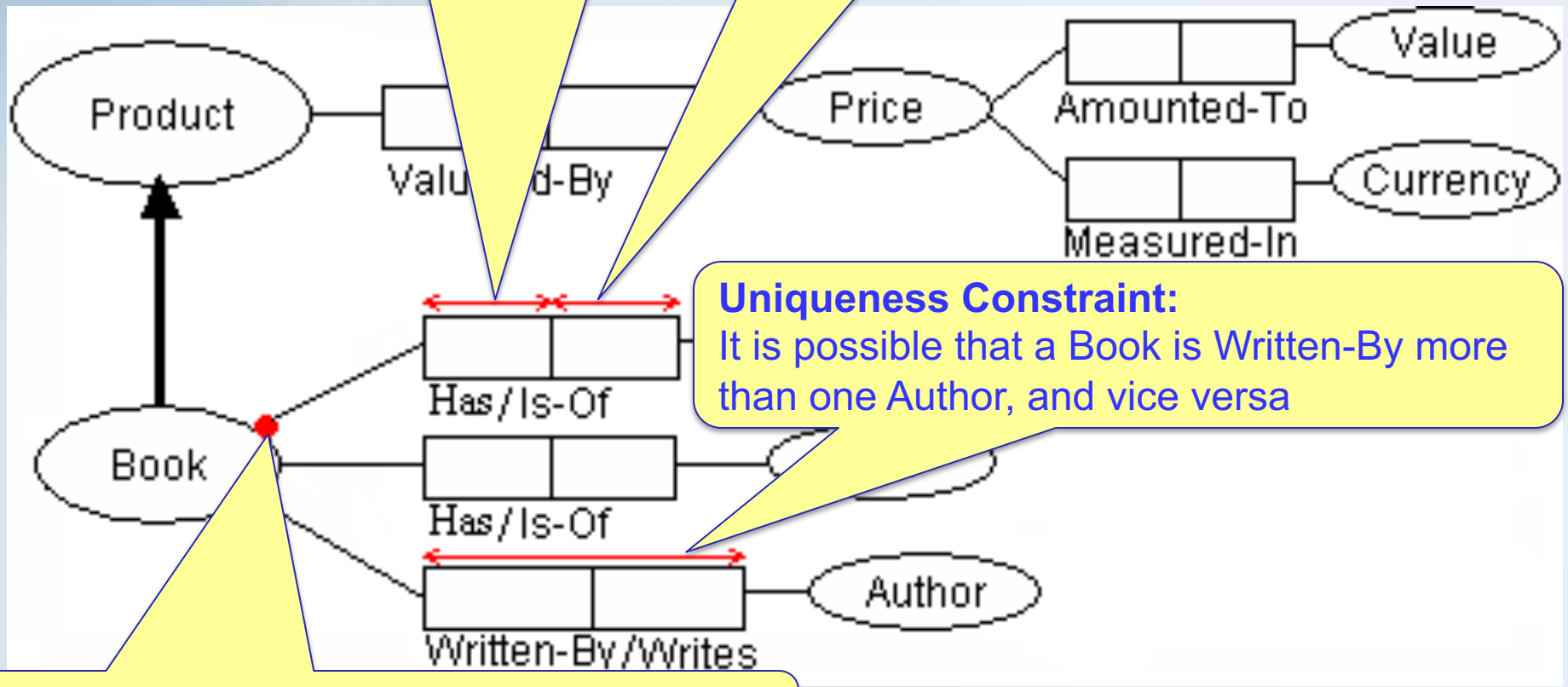
Object-Role Modeling (ORM) constructs



Object-Role Modeling (ORM) constraints

Uniqueness Constraint:
Each ISBN must Is-Of at most one Book

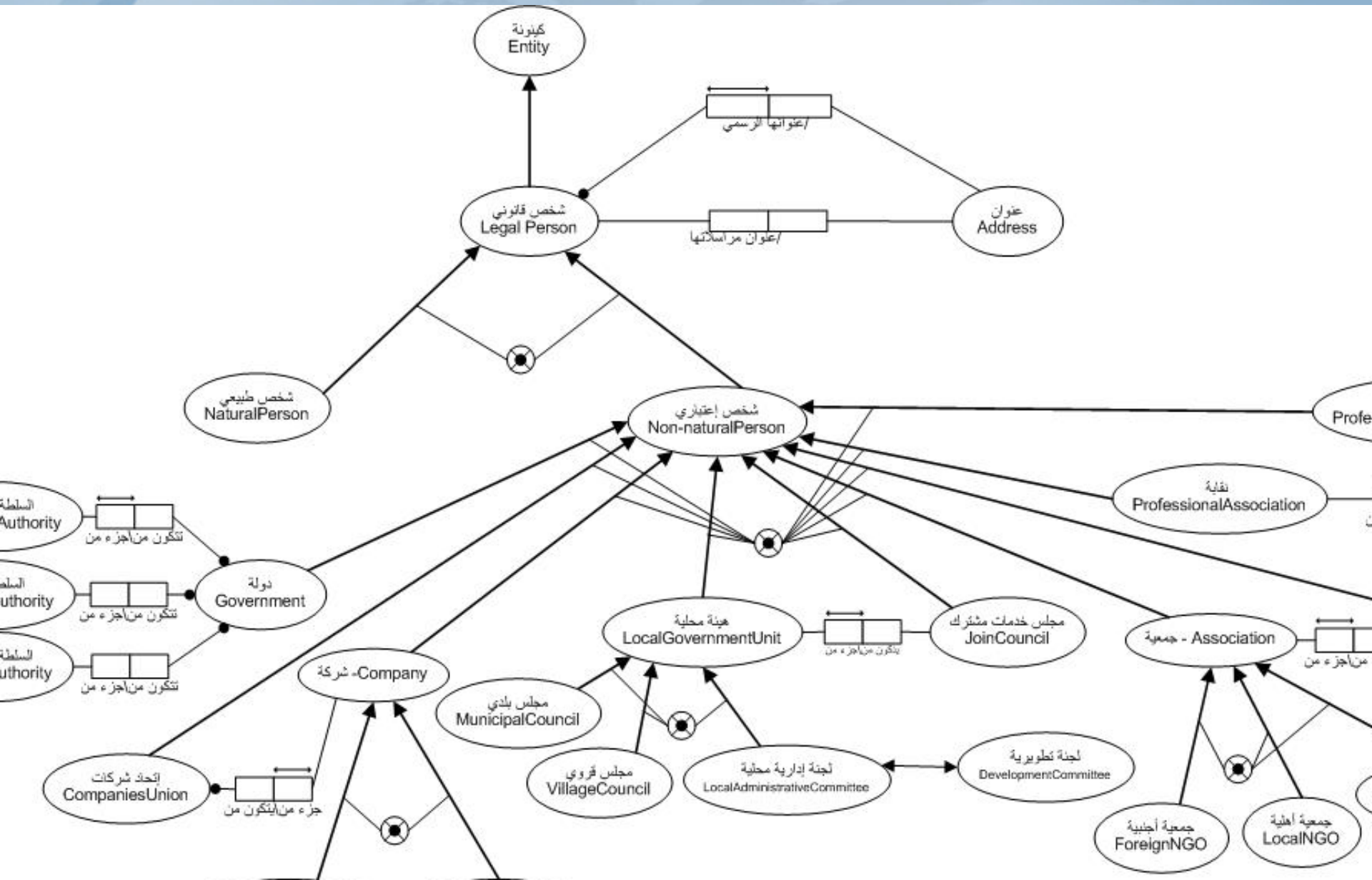
Uniqueness Constraint:
Each Book must has at most one ISBN



Uniqueness Constraint:
It is possible that a Book is Written-By more than one Author, and vice versa

Mandatory Constraint:
Each Book must has at Least one ISBN

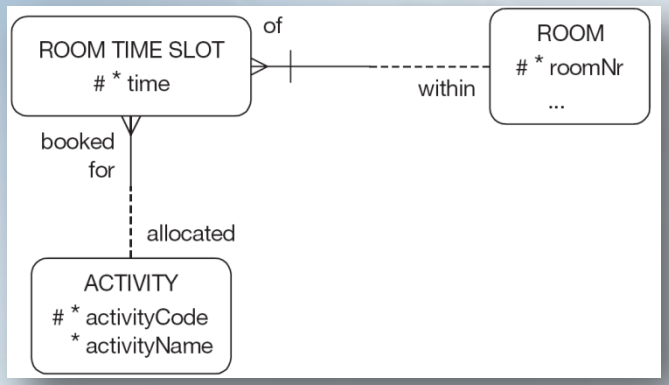
eGov Ontology (using ORM)



Modeling Approaches (short discussion)

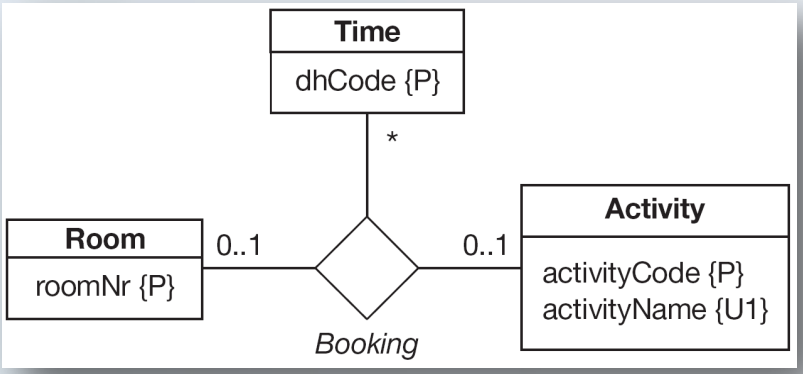
Which is more intuitive for modelers? For domain experts?

ER-model



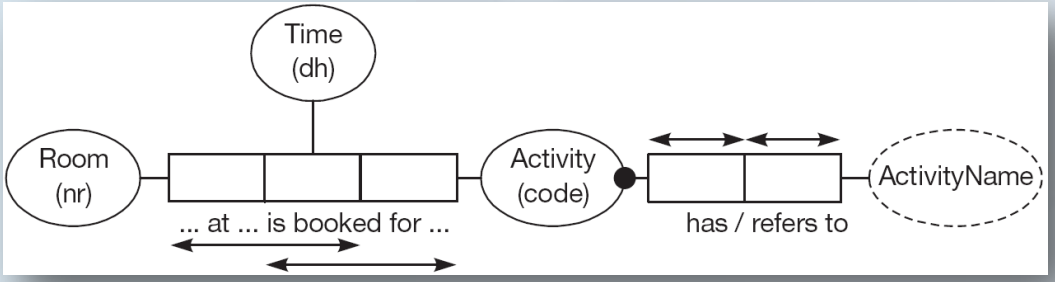
Nice for picturing to DB-schemes

UML-model



Close to the way programmers think

ORM-model



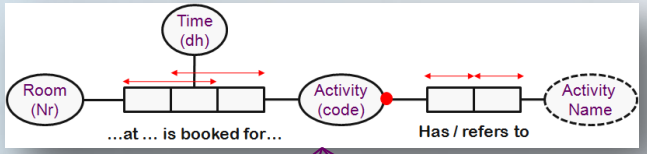
Suitable for general conceptual modeling, not only DB schemes

Outline

- Information Modeling
- Modeling Approaches
- Introduction to Object Role Modeling (ORM)
- Information Levels

Information Levels (Data Modeling Viewpoint)

Conceptual Level



- What kind of facts/concepts we need, and how they are related.
- Conceptual models are designed for clear communication, especially between modelers and domain experts.

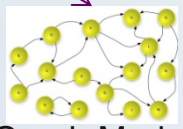
Logical Level

Relational
DB

Object Oriented
OO-DB

...

Tree model
XML



Graph Model
RDF

- Abstract data structures
- Same conceptual schema can be mapped into several logical structures

Physical Level

...

- The physical storage and access structures used in a system (indexes, file clustering, etc.).
- Same Logical schema can be stored in different ways

Information Levels (Data Modeling Viewpoint)

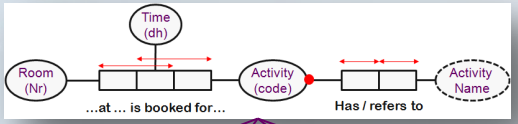
Linguistic Level

- Concerned with the terms used to lexicalize the meaning.
- Same meaning can be lexicalized in deferent languages.

Ontological Level

- Concerned with the meaning, in the real world.
- Same meaning (/intentions) can be conceptualized in different ways.

Conceptual Level



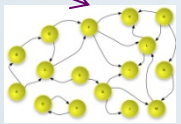
- What kind of facts/concepts we need, and how they are related.
- Conceptual models are designed for clear communication, especially between modelers and domain experts.

Logical Level

Relational
DB

Object Oriented
OO-DB

Tree model
XML



Graph Model
RDF

- Abstract data structures
- Same conceptual schema can be mapped into several logical structures

Physical Level

- The physical storage and access structures used in a system (indexes, file clustering, etc.).
- Same Logical schema can be stored in different ways

Knowledge Levels (from philosophy viewpoint)

[Guarino]

Level	*Primitives*	*Interpretation*	*Main feature*
Linguistic	Linguistic terms	Subjective	Language dependence
Conceptual	Conceptual relations	Subjective	Conceptualization
Ontological	**Ontological relations**	**Constrained**	**Meaning**
Epistemological	Structuring relations	Arbitrary	Structure
Logical	Predicates, functions	Arbitrary	Formalization

- Will be discussed later

Information Levels (Data Modeling Viewpoint)

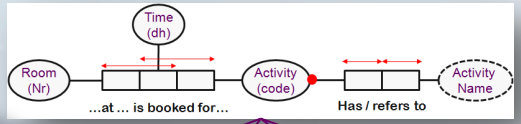
Linguistic Level

- Concerned with the terms used to lexicalize the meaning.
- Same meaning can be lexicalized in deferent languages.

Ontological Level

- Concerned with the meaning, in the real world.
- Same meaning (/intentions) can be conceptualized in different ways.

Conceptual Level



- What kind of facts/concepts we need, and how they are related.
- Conceptual models are designed for clear communication, especially between modelers and domain experts.

Logical Level

- Abstract data structures
- Same conceptual schema

Physical Level

- ORM is the most suitable language for conceptual modeling (not only conceptual data modeling). That is, it allows modelers to think more conceptually and be more independent from the logical level.
- ORM is also being used as ontology modeling language, business rules and requirements specification, XML-schema modeling, etc. (not only DB modeling)

ORM Tools (feel free to use any tool in the course)

➔Microsoft Visio

➔Other Tools:

- **Microsoft VisioModeler (**Free but not supported anymore)
  - downloadable from http://www.cnet.com.au/downloads/soa/Visio-2000-Tool-VisioModeler-Unsupported-Product-Edition/0,239030384,10626416s,00.htm
  - you need replace a DLL to run it in Win7.
- **NORMA**
  - downloadable from SourceForge or http://www.ormfoundation.org/files/folders/norma_the_software/default.aspx.
  - Free and open source (but you need Visual Studio 2005 or 2008 to run it).
  - Supports ORM2
- **DogmaModeler**
  - downloadable from http://www.jarrar.info/Dogmamodeler/
  - Free and open source (prototype status)
  - Designed as Ontology modeling tool (Norma and VisioModeler are database tools)
  - Will be required later in the course.

References

[1] Terry Halpin, Tony Morgan: Information Modeling and Relational Databases, Second Edition. Second Edition. The Morgan Kaufmann Series in Data Management Systems. ISBN: 0123735688

[2] Mustafa Jarrar and Robert Meersman: Ontology Engineering -The DOGMA Approach. Book Chapter in "Advances in Web Semantics I". Chapter 3. Pages 7-34. LNCS 4891, Springer.ISBN:978-3540897835. (2008).

[3] Mustafa Jarrar, Anton Deik, Bilal Faraj: Ontology-Based Data And Process Governance Framework -The Case Of E-Government Interoperability In Palestine . In pre-proceedings of the IFIP International Symposium on Data-Driven Process Discovery and Analysis (SIMPDA'11). Pages(83-98). ISBN 978-88-903120-2-1. Campione, Italy. June 30, 2011.

[4] Mustafa Jarrar: Mapping ORM Into The SHOIN/OWL Description Logic- Towards A Methodological And Expressive Graphical Notation For Ontology Engineering . In OTM 2007 workshops: Proceedings of the International Workshop on Object-Role Modeling (ORM'07). Pages (729-741), LNCS 4805, Springer. ISBN: 9783540768890. Portogal. November, 2007

[5] Mustafa Jarrar: Towards Automated Reasoning On ORM Schemes. -Mapping ORM Into The DLR_idf Description Logic. In proceedings of the 26th International Conference on Conceptual Modeling (ER 2007). Pages (181-197). LNCS 4801, Springer. Auckland, New Zealand. ISBN 9783540755623. November 2007

[6] Mustafa Jarrar and Stijn Heymans: Unsatisfiability Reasoning In ORM Conceptual Schemes. In Current Trends in Database Technology - EDBT 2006: Proceeding of the IFIP-2.6 International Conference on Semantics of a Networked. Pages (517-534). LNCS 4254, Springer. Munich, Germany. ISBN: 3540467882. March 2006.

[7] Mustafa Jarrar and Stijn Heymans: Towards Pattern-Based Reasoning For Friendly Ontology Debugging . Journal of Artificial Intelligence Tools. Volume 17. No.4. World Scientific Publishing. August 2008.

[8] Mustafa Jarrar, Maria Keet, and Paolo Dongilli: Multilingual Verbalization Of ORM Conceptual Models And Axiomatized Ontologies. Technical report. STARLab, Vrije Universiteit Brussel, February 2006.

[9] Sergey Lukichev and Mustafa Jarrar: Graphical Notations For Rule Modeling . Book chapter in "Handbook of Research on Emerging Rule-Based Languages and Technologies". IGI Global. ISBN:1-60566-402-2. (2009)

[10] Mustafa Jarrar: Modularization And Automatic Composition Of Object-Role Modeling (ORM) Schemes .OTM 2005 Workshops: Proceedings of the Object-Role Modeling (ORM'05). Pages (613-625). LNCS 3762, Springer. ISBN: 3540297391. 2005.

[11] Mustafa Jarrar: Towards Methodological Principles For Ontology Engineering. PhD Thesis. Vrije Universiteit Brussel. (May 2005)

[12] Mustafa Jarrar, Jan Demey, and Robert Meersman: On Using Conceptual Data Modeling For Ontology Engineering . Journal on Data Semantics, Special issue on "Best papers from the ER/ODBASE/COOPIS 2002 Conferences". LNCS 2800. No 1. Springer. 2003.

[13] Jan Demey, Mustafa Jarrar, and Robert Meersman: A Markup Language For ORM Business Rules . Proceedings of the International Workshop on Rule Markup Languages for Business Rules on the Semantic Web (RuleML 2002). Pages(107-128). Volume 60. CEUR Workshop Proceedings. ISSN 1613-0073. June 2002

[14] Mustafa Jarrar: Towards Effectiveness And Transparency In E-Business Transactions, An Ontology For Customer Complaint Management . A book chapter in "Semantic Web Methodologies for E-Business Applications". chapter 7. IGI Global. (2008)

[15] Mustafa Jarrar: ORM Markup Language, Version 3 . Technical Report. STAR Lab, Vrije Universiteit Brussel, Belgium. January 2007