12th Workshop for PhD Students in Object-Oriented Systems (PhDOOS 2002), 16th European Conference on Object-Oriented Programming (ECOOP 2002), p. 48-49: Lecture Notes in Computer Science 2548, Málaga (Spain), June 10-14 2002.

Departamento de Lenguajes y Sistemas Informáticos

Universitat d'Alacant
Universidad de Alicante

Multidimensional Modeling using UML and XML

Sergio Luján-Mora

Contents

Introduction

- Multidimensional (MD) modeling → Data warehouses, MD databases, OLAP applications
- Many years of historical information
- Different approaches for the conceptual design:
  - Golfarelli *et al*
  - Sapia *et al*
  - Tryfona *et al*
  - …

Own graphical notations

Multidimensional Modeling using UML and XML

Introduction

- UML → Standard OO modeling language for software systems
- Minimize the efforts in learning new notations
- Extensible language → Stereotypes, tagged values, and constraints
- Allow introducing new elements for specific domains

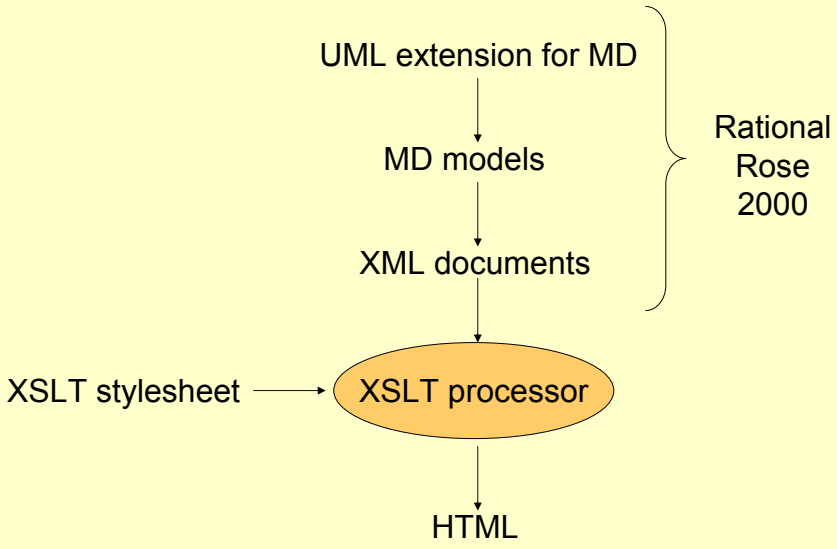
Introduction

- UML extension for MD modeling based on an object-oriented approach (Trujillo *et al*, *IEEE Computer 34, 2001*):
  - Easily considers MD properties at the conceptual level:
    - Many-to-many relationships
    - Degenerate dimensions
    - Multiple and alternative path hierarchies
    - ...

Introduction

- MD models are stored in XML documents → XML Schema defines the correctness
- Then, we use XSLT to automatically generate HTML pages that can represent different presentations of the same MD model

Introduction



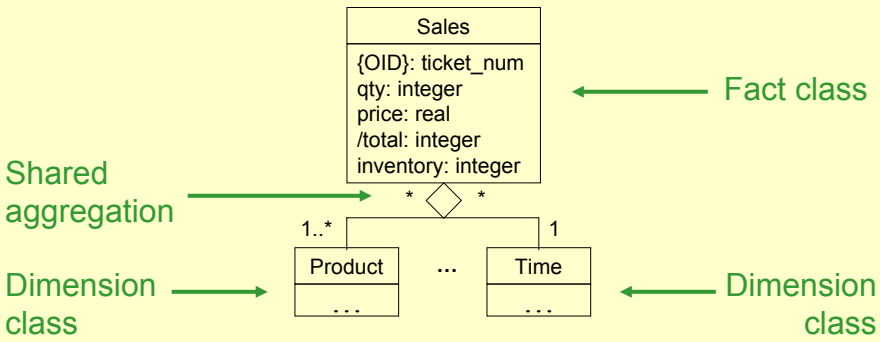
Contents

- Introduction
- **OO Multidimensional Modeling**
- UML Extension for MD Modeling
- MD Modeling in Rational Rose
- MD Models in XML
- Conclusions and Future Work

OO MD modeling

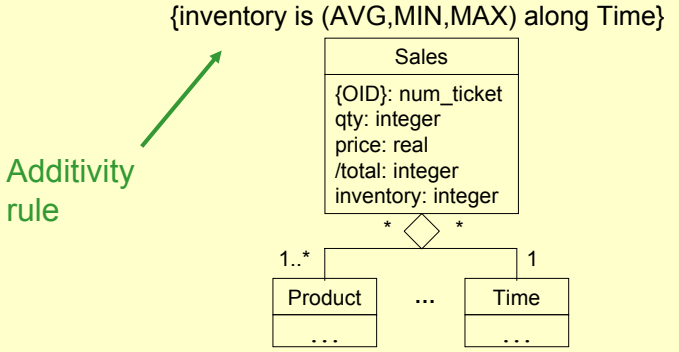
- The MD modeling approach represents both the structural and ~~dynamic~~ parts of MD modeling using the UML
- MD modeling structural properties are specified by means of a UML class diagram
- Facts and dimensions are considered by *fact classes* and *dimension classes*

OO MD modeling



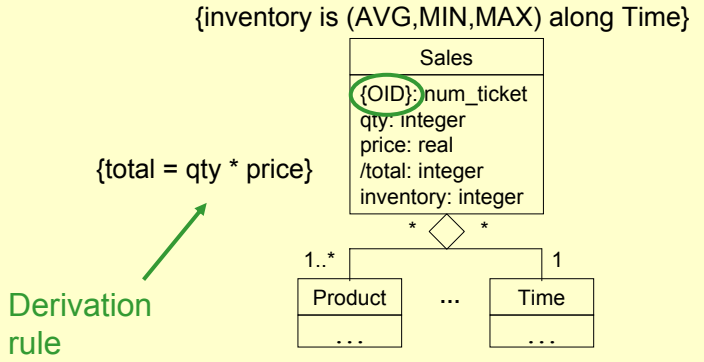
•Fact classes are specified as composite classes in shared aggregation relationships of *n* dimension classes

OO MD modeling



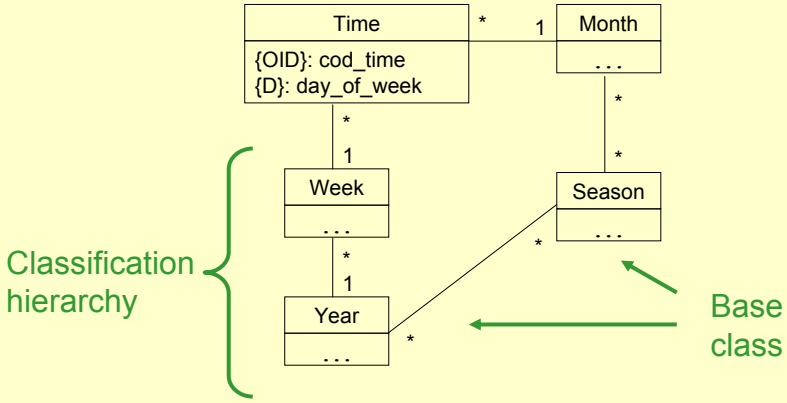
- All measures in the fact class are considered additive
- Non-additive measures → *Additivity rules* defined as constraints

OO MD modeling



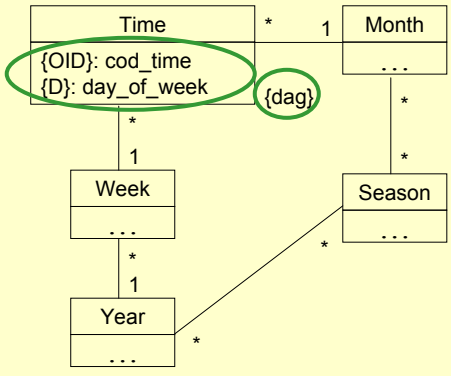
- Derived measures are defined by means of *derivation rules*
- *Identifying attributes* **{OID}** → Represent degenerate dimensions

OO MD modeling



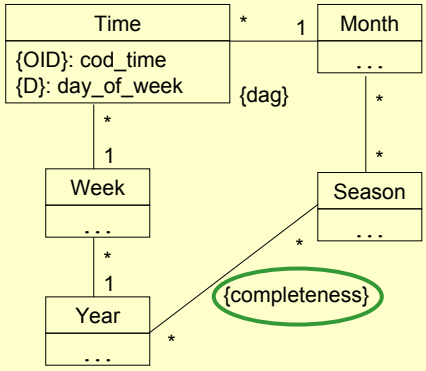
- An association of classes specifies the relationships between two levels of a **classification hierarchy**
- Every classification hierarchy level is specified by a class called **base class**

OO MD modeling



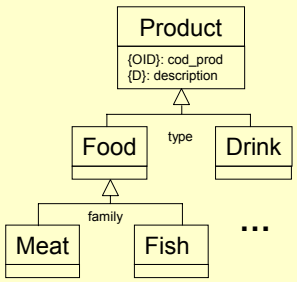
- The classes in a classification hierarchy must define a Directed Acyclic Graph (DAG) rooted in the dimension class (**{dag}**)
- Every classification hierarchy level must have an *identifying attribute* (**{OID}**) and a *descriptor attribute* (**{D}**)

OO MD modeling



- The multiplicity **1** and **1..*** addresses the concepts of *strictness* and *non-strictness*
- The **{completeness}** constraint addresses the *completeness* of a classification hierarchy

OO MD modeling



- The categorization of dimensions is considered by means of generalization-specialization relationships

Contents

UML Extension for MD Modeling

- UML extensible language → Extension mechanisms: stereotypes, tagged values, and constraints
- Allow introducing new elements for specific domains (web design, data modeling, etc.)
- UML can be adapted to fit a specific method, organization, or user

UML Extension for MD Modeling

- **Stereotype**: a new model element that specializes a UML element (Class, Attribute, Package, Association, etc.)
- **Tagged value**: a new property of a model element
- **Constraint**: refines the semantics of a model element → Informal or formal (Object Constraint Language)

UML Extension for MD Modeling

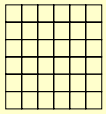
- Extension summary:
  - 8 stereotypes:
    - Class: **Fact**, **Dimension**, and **Base**
    - Attribute: **FactAttribute**, **DimensionAttribute**, **OID**, and **Descriptor**
    - Association: **Completeness**
  - 2 tagged values:
    - **isTime** and **derivationRule**
  - 23 constraints

UML Extension for MD Modeling

- Facts and dimensions → **Fact** and **Dimension** stereotypes
- Derived measures: **derivationRule** tagged value
- Classification hierarchies → Association between **Dimension** and **Base** stereotypes
- Completeness → **Completeness** stereotype

UML Extension for MD Modeling

- Name: **Fact**
- Base class: **Class**
- Description: **Classes of this stereotype represent facts in a MD model**
- Icon:



- Tagged values: **None**

UML Extension for MD Modeling

- Constraints:
  - All attributes of a Fact must be OID or FactAttribute:

    self.feature->select(oclIsKindOf(Attribute))->
    forAll(oclIsTypeOf(OID) or oclIsTypeOf(FactAttribute))
  - All associations of a Fact must be aggregations:

    self.association->forAll(aggregation = #aggregate)
  - A Fact can only be associated to Dimension classes:

    self.allOppositeAssociationEnds->
    forAll(participant.oclIsTypeOf(Dimension))

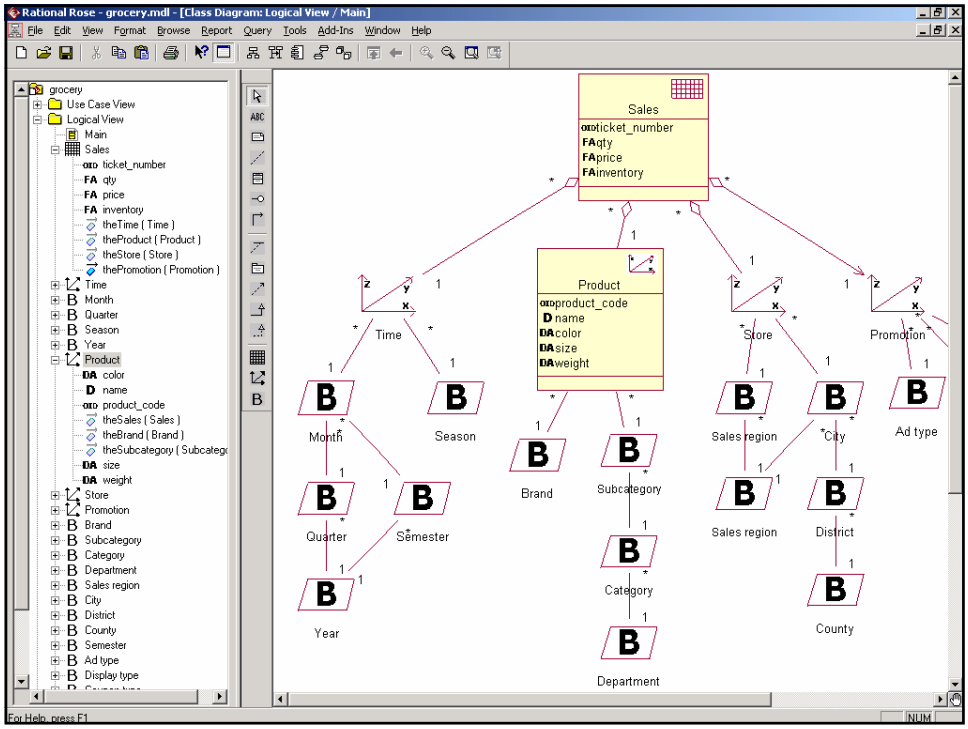
Contents

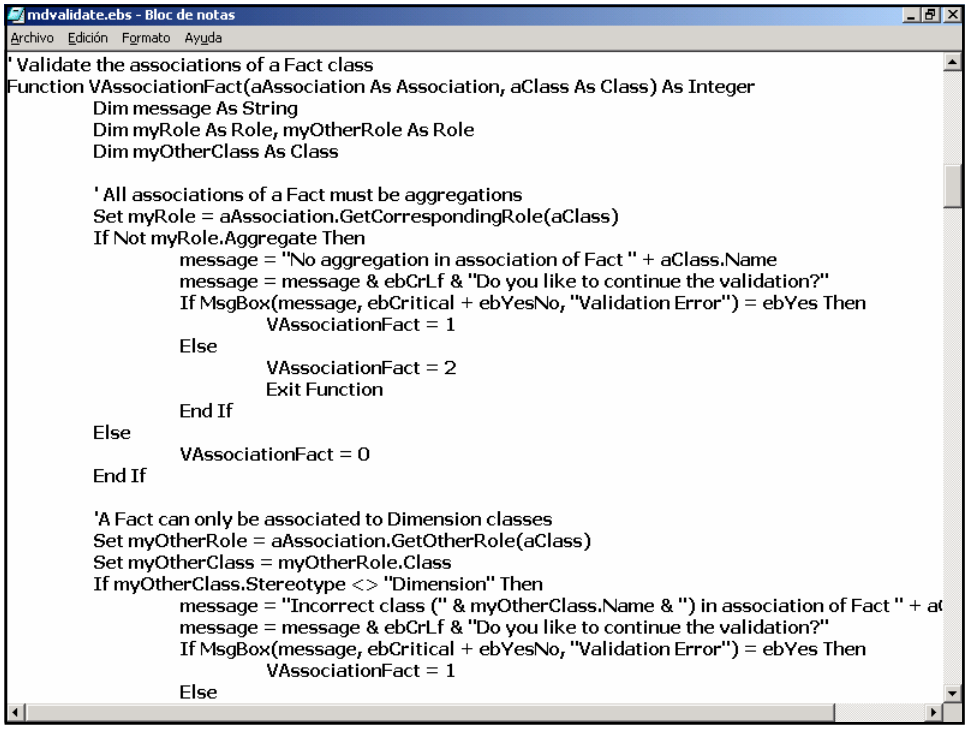
MD Modeling in Rational Rose

- Rational Rose is one of the most well-known visual modeling tools
- RR is extensible by means of add-ins through the Rose Extensibility Interface:
  - Main menu items
  - Stereotypes
  - Properties (*tagged values*)
  - Data types
  - Event handling
  - Scripts
  - …

MD Modeling in Rational Rose

- Our add-in customizes:
  - Stereotypes → Stereotype configuration file
  - Properties → Property configuration file
  - Menu item → Menu configuration file
    - Menu Tools:
      - MD Validate
      - XML Export


Rational Rose - grocery.mdl - [Class Diagram: Logical View / Main]
File Edit View Format Browse Report Query Tools Add-Ins Window Help
grocery
Use Case View
Logical View
Main
Sales
ticket_number
qty
price
inventory
theTime (Time)
theProduct (Product)
theStore (Store)
thePromotion (Promotion)
Time
Month
Quarter
Season
Year
Product
color
name
product_code
theSales (Sales)
theBrand (Brand)
theSubcategory (Subcateg
size
weight
Store
Promotion
Brand
Subcategory
Category
Department
Sales region
City
District
County
Semester
Ad type
Display type
Sales
ticket_number
qty
price
inventory
Product
product_code
name
color
size
weight
Time
Store
Promotion
Month
Season
Brand
Subcategory
Sales region
City
Ad type
Quarter
Semester
Sales region
District
Year
Category
County
Department
For Help, press F1
NUM


mdvalidate.ebs - Bloc de notas
Archivo Edición Formato Ayuda
' Validate the associations of a Fact class
Function VAssociationFact(aAssociation As Association, aClass As Class) As Integer
Dim message As String
Dim myRole As Role, myOtherRole As Role
Dim myOtherClass As Class
' All associations of a Fact must be aggregations
Set myRole = aAssociation.GetCorrespondingRole(aClass)
If Not myRole.Aggregate Then
message = "No aggregation in association of Fact " + aClass.Name
message = message & ebCrLf & "Do you like to continue the validation?"
If MsgBox(message, ebCritical + ebYesNo, "Validation Error") = ebYes Then
VAssociationFact = 1
Else
VAssociationFact = 2
Exit Function
End If
Else
VAssociationFact = 0
End If
'A Fact can only be associated to Dimension classes
Set myOtherRole = aAssociation.GetOtherRole(aClass)
Set myOtherClass = myOtherRole.Class
If myOtherClass.Stereotype <> "Dimension" Then
message = "Incorrect class (" & myOtherClass.Name & ") in association of Fact " + a
message = message & ebCrLf & "Do you like to continue the validation?"
If MsgBox(message, ebCritical + ebYesNo, "Validation Error") = ebYes Then
VAssociationFact = 1
Else

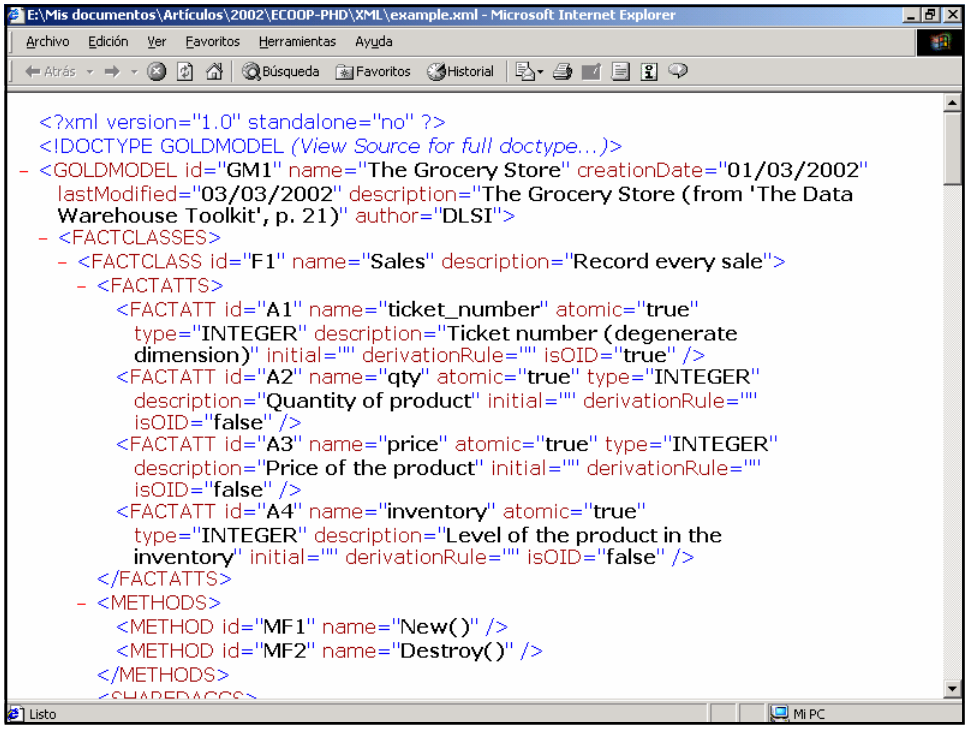
Contents

MD Models in XML

- XML is being adopted as a standard syntax for the interchange of semi-structured data
- We use XML to store MD models
- Correct structure? → XML Schema

MD Models in XML

- Main advantages of XML Schema over DTD:
  - They are written in the same syntax as XML documents
  - They can define new data types
  - The references are more precisely defined
- We have chosen a “Russian doll” design (nested, anonymous complex types)



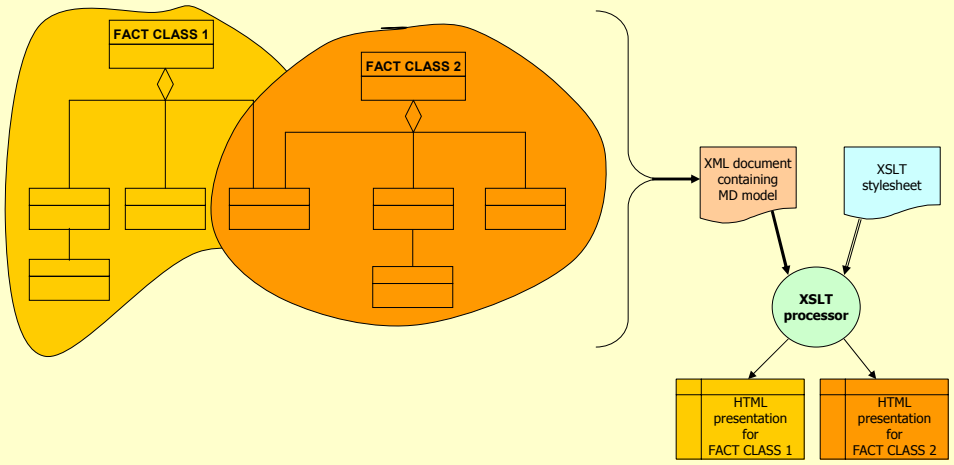
MD Models in XML

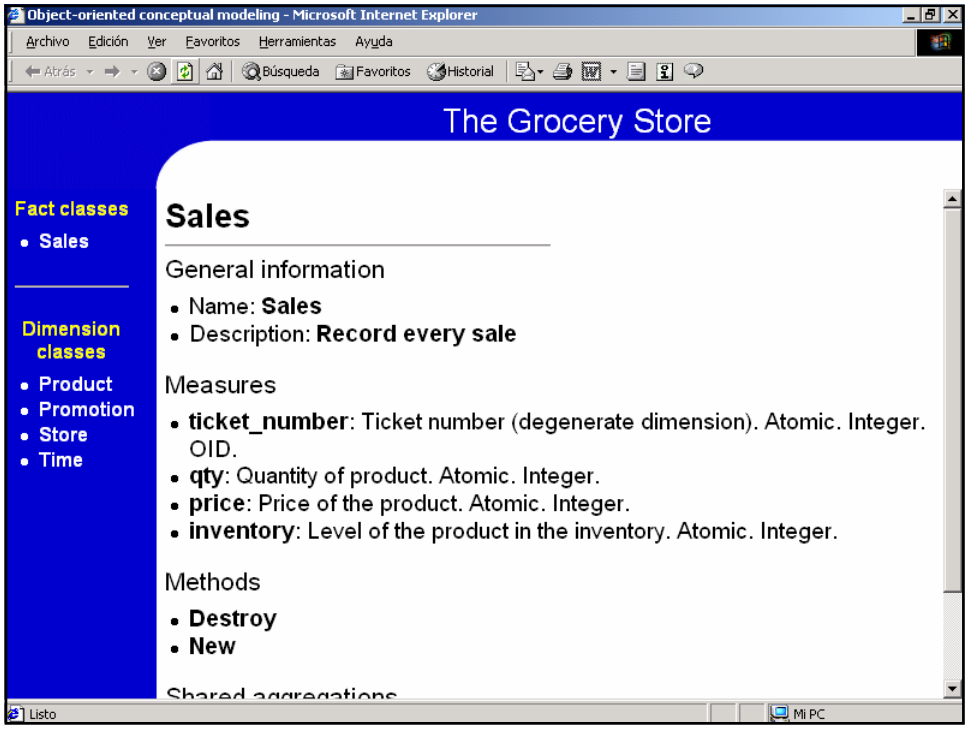
- Goal: provide different presentations of MD models in the web
- Common web browsers partly support XML
- We are currently forced to transform XML documents into HTML pages in order to publish them in the web
- How?

MD Models in XML

- XSLT is the best method: it is a language for transforming XML documents into other XML documents (XML → XHTML)
- XML documents can be tailored (filtered and reordered) to represent different presentations of the same MD model using XSLT stylesheets

MD Models in XML





Contents

Conclusions and Future Work

- UML extension for MD modeling: allows us to represent structural MD properties at the conceptual level
- OCL to specify the constraints, avoiding an arbitrary use of the extension
- Main advantage: UML → Avoids developers learning a new graphical notation

Conclusions and Future Work

- MD models are stored in XML documents → We provide an XML Schema
- Different presentations from the same MD model in HTML → We provide XSLT stylesheets

Multidimensional Modeling using UML and XML

Conclusions and Future Work

- PhD: define a methodology for MD modeling
- Until now: graphical notation, static part, representation in XML, some design guidelines
- Future work: dynamic part, UML package diagrams, automatic generation of database schema into OO and OR databases, more design guidelines, …

Departamento de Lenguajes y Sistemas Informáticos

Universitat d'Alacant
Universidad de Alicante

Multidimensional Modeling using UML and XML

Sergio Luján-Mora